DOCUMENT RESUME

ED 426 750 JC 990 084

AUTHOR Armstrong, William B.
TITLE Explaining Community College Outcomes by Analyzing Student Data and Instructor Effects.
PUB DATE 1999-00-00
NOTE 246p.; Ph.D. Dissertation, University of California, Los Angeles.
PUB TYPE Dissertations/Theses - Doctoral Dissertations (041)
EDRS PRICE MF01/PC10 Plus Postage.
DESCRIPTORS *Academic Achievement; *Community Colleges; High Schools; *Predictor Variables; *Student Characteristics; Tables (Data); Test Validity; *Testing; Two Year College Students; Two Year Colleges
IDENTIFIERS Placement Tests

ABSTRACT

This study examines the validity and usefulness of employing standardized placement tests and other indicators of academic ability to make education or training decisions about community college students. It also demonstrates how attempts to make optimal educational placement decisions through mandatory placement testing are confounded by measurement error in tests, differing characteristics of students, and the instability of the criterion variables of grade or retention. The investigation was conducted in four parts. Part one focused on descriptive data about the participant sample; part two examined the validity of standardized placement test scores in predicting student course outcomes; part three explored the relation of student demographic, dispositional, and situational variables to course outcomes; and part four examined the role of the instructor as a source of variance in explaining or predicting course outcomes. Chapters include a review of historical and sociological foundations of current debates over testing in the community college; a review of the literature; the study's methodology; findings; and conclusions and recommendations. In general, the study revealed that dispositional variables, such as high school grade point average, educational goal, and importance of college attendance to respondent, tended to account for greater variance in course outcomes of retention and final grade than did other biographical data, including placement test scores. Appended are placement test scores, significant variables, and a copy of the questionnaire. (Contains 52 tables and 166 references.) (EMH)

UNIVERSITY OF CALIFORNIA

Los Angeles

Explaining Community College Outcomes by Analyzing Student Data and Instructor Effects

A dissertation submitted in partial fulfillment of the requirements for the degree

Doctor of Philosophy in Education

by

William Bannon Armstrong

1999

The dissertation of William Bannon Armstrong is approved:

Marvin C. Alkin

Jeannie Oakes

Leobardo F. Estrada

Arthur M. Cohen, Committee Chair

University of California, Los Angeles

1999

DEDICATION

There are so many people that affect the course of your life, it is difficult to list them all. Perhaps first I must dedicate this dissertation to my mother, Dorothy Ellen Armstrong, and father, George Bannon Armstrong. I thank my mother for bestowing on me her gifts of compassion, empathy, and humor. To my father I am grateful for his lessons on perseverance, a positive attitude, an indefatigable spirit, and the ability to make others laugh. I thank my brother Doug, whose probing and insistent questions as to my progress on this project helped to keep me focused. If one is a reflection of the family that raises them, then I am truly fortunate.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to gratefully acknowledge the members of my dissertation committee whom I selected on the basis of my strong admiration for their work, and also for the affect they had on me as a scholar. I want to give special thanks to my adviser and committee chair, Dr. Arthur M. Cohen, for his patience and encouragement at every stage of the graduate school and dissertation process. Dr. Cohen believed in the value of this project and shared his insights and knowledge to the research questions. He has helped to mold and shape my scholarship in education. His perspective and insights into the open-access college shed light on many facets of this study. I am indeed fortunate to have worked with him as a student, author, and graduate assistant. His scholarship will forever be an inspiration to me.

I want to thank Dr. Marvin Alkin for his insights into evaluation and his faith in this project. Dr. Alkin's background and professional experience with the problems posed by this investigation were of enormous value in framing the study. His encouragement of the project was a great help. His ability to engage students in the process of evaluation and inquiry is unique. I have benefited immensely from our association.

My appreciation also goes to Dr. Jeannie Oakes who provided me with her perspectives on student ability and school reform both in her writing and in class. Dr. Oakes' writings and course lectures helped to inspire the topic of this investigation and several lines of the research followed in this study. I will never forget her admonition in class that we conduct research on educational practice with the overall goal of making things better. I hope in some way that this study can help to inform policy and influence practice in the community colleges to indeed make things better.

Dr. Leobardo Estrada was steadfast in his support of this project and for this I am grateful. Dr. Estrada always made himself available to me, and given the demands on his

time by students, this was no easy feat. He has a core belief in the value of the open-access college as a vital institution for the economic and social betterment of the community. His perspectives on community college educational policy provided important guidance for the public policy implications of this research. Dr. Estrada demonstrated to me his deep commitment to students and scholarship. He is a man of integrity whose actions reflect his words. He always asked the tough questions, and he contributed greatly to this project.

I tested many a friendship through this project. My friends helped to carry me through the difficult times with their words of encouragement, hope, and occasional scolding. I want to thank my best friend and colleague. Dr. William P. Evans for his being there when I needed him, and for helping me throughout this process. His sense of humor and genuinely human perspective on life was a great source of help in this project. Richard and Carolyn Hofstetter were tireless in their support of my project. They provided me with invaluable encouragement. To them I owe a debt of gratitude. Dr. Judy Stamm for her tireless phone calls about my status, and the personal interest she took in my completing this project. No one knows the journey you are on lest they walk it themselves. Judy knew the journey, and helped me throughout both as a friend and confidante. I also want to thank Dr. Troni Rifkin and the staff at the ERIC Clearinghouse for Community Colleges for their invaluable assistance.

I also want to thank my friend and mentor, Dr. Thomas G. Sticht. When we set out to write a book, it was Tom who first convinced me that I could write and conduct research. He stuck with me and helped to train me in the research process. I owe much of my professional development to Tom. I will never forget that.

I wish to also acknowledge the contributions of Augie Gallego, Chancellor of the San Diego Community College District. Augie demonstrated his commitment to the value of education and helped enable me to pursue graduate studies. He practices what he values, and for that I am fortunate.

VITA

September 24, 1956	Born, Ft. Lauderdale, Florida
1979	B.A. Liberal Studies San Diego State University San Diego, California
1980-1982	Graduate Studies, Political Science, San Diego State University
1982-1984	Teaching Assistant, Educational Opportunity Program San Diego State University
1984-1986	Project Director, Applied Behavioral & Cognitive Sciences, Inc. San Diego, CA
1986-1988	Comprehensive Adult Student Assessment System San Diego, CA
1990	Adjunct Faculty, San Diego City College
1988-1990	Institutional Research Analyst San Diego Community College District
1991-1993	Graduate Student Assistant Center for the Study of Community Colleges Los Angeles, CA
1992	Reviewer, UCLA Journal of Education
1993	Reviewer, Journal for Research in Mathematics Education
1994	Project Principal Investigator: Compendium of Adult Literacy. Project funded by the National Institute for Literacy
1995	Achieved Honors on Doctoral Qualifying Exam
1995	Appointed to Joint Commission for Accountability Reporting, Washington, DC
1996	Project Director: Content Standards in Adult Basic Education (CONSABE) Project. Grant awarded by National Institute for Literacy
1991-Present	Director of Institutional Research, San Diego Community College District

PUBLICATIONS AND PRESENTATIONS

Sticht, Armstrong, Hickey, and Caylor (1987). Cast-off youth: Policy and training methods from the military experience. New York: Praeger Publishing

Armstrong, William B.(1993, Spring). Accountability and a unitary transfer definition: Report of the 1992 Transfer Assembly. Community College Review: 5 59-68

Armstrong, William B. and Mellisinos, M (1994, Summer). Examining the relationship between liberal Arts, course levels, and transfer rates. New Directions for Community Colleges, 2 . 81-91.

Sticht., T.G., and W. Armstrong (1996). Adult literacy in the United States: A compendium of quantitative data and interpretive comments. Washington, DC: National Institute for Literacy.

Spicer and Armstrong (1997). Transfer: The elusive denominator. New Directions for Community Colleges . 96 45-54.

Assessing the Campus Climate of the Community College: Student, Faculty, and Staff Perceptions. Paper presented at the Annual meeting of the American Educational Research Association (San Diego, CA: March 1998).

Validating Placement Tests for Course Performance. Paper presented at the annual meeting of the American Evaluation Association. (San Diego, CA, 1997).

The Relationship of Perceptions of the Campus Climate to Student Outcomes. Paper presented at the Annual Forum of the Association for Institutional Research (36th, Albuquerque, NM: May 1996).

Building Indicators of Transfer Effectiveness and Student Equity in the San Diego Community College District: A Local Application of the Transfer Assembly Model. Paper presented at the Association for the Study of Higher Education Annual Conference. 19th, Tucson, AZ: November 1994.

ABSTRACT OF THE DISSERTATION

Explaining Community College Course Outcomes by

Analyzing Student Data and Instructor Effects

by

William Bannon Armstrong

Doctor of Philosophy in Education

University of California, Los Angeles, 1999

Professor Arthur M. Cohen, Chair

In the last ten years, there has been rapid growth in the use of standardized tests to admit and place students into community college courses. Growth in the use of testing has prompted debate over the equity, access, and fairness implications of placement testing and mandatory pre-requisites for course eligibility. This has resulted in state regulations requiring that California community colleges provide validity evidence of test score-based inferences and course pre-requisites. However for both theoretical and technical reasons, the observed predictive validity coefficients between placement test scores and final grades or retention in a course often demonstrate a weak relationship. Thus if a placement test fails to predict an outcome, or if students are misclassified as unlikely to succeed in a course, we may not know if this is due to the test, the characteristics of students, or the instability of the criterion variable. A review of predictive validity studies in higher education showed

that many did not include student biographical information as a method to improve the accuracy and fairness of placement practices. Also, few studies attempted to account for differential grading practices or teacher characteristics as a potential source of error in predicting student course performance. After a review of several measurement theories, this study applied the principles of Point-to-Point and Aptitude-Treatment Interaction theories to model and identify the extent to which test score, biographical, situational, dispositional, and instructor characteristics explain variance in college course outcomes. This analysis revealed dispositional variables (high school grade point average, grades received in last English or mathematics course, educational goal, importance of college attendance to respondent, years of high school English, and years since last mathematics course) tended to account for greater variance in course outcomes of retention and final grade than did other biographical data, including placement test scores. Instructor grading variation significantly reduced error in the model indicating that grading inconsistency or differences in retention practices has a strong effect on student course performance. In general, standards reflect less what the student brings to the college than what the institution imposes on the student.

CHAPTER ONE

INTRODUCTION

This study is about the use of placement testing to make inferences about the potential academic success of students in the community colleges. The problem I focus on in this dissertation is the defensibility of test score-based placement or academic prerequisite requirements used to predict performance in community college courses. Prediction of student performance or retention in courses is often confounded when the criterion variable is unstable or influenced by other variables such as student and instructor characteristics. This investigation examines the validity and usefulness of employing standardized placement tests and other indicators of academic ability to make education or training decisions about the person being tested. This investigation also demonstrates how attempts to make optimal educational placement decisions through mandatory placement testing are confounded by measurement error in tests, differing characteristics of students, or the instability of the criterion variable of grade or retention.

The open-access mission of the community college practically ensures that the students served will differ tremendously in their experiences, levels of education, and socio-economic status. Predicting success is made more difficult by the diversity in the dispositional and situational characteristics of students. In addition, poor predictions of success or retention may occur because teachers --acting as individual raters-- differ in their evaluation practices and approaches to instruction.

This study encompasses several issues regarding faculty, students, instruction, and institutional policies. Community college reform legislation passed in the mid-1980's mandated that 75% of courses were to be taught by full-time, professional faculty. The rationale was that using more full-time, contract faculty would improve the quality of

instruction in the community colleges. Greater reliance on full-time faculty was to improve instructional outcomes in the colleges because of their closer connection to the college. Faculty connection to the college was enhanced through regular attendance at faculty meetings, offices on campus, greater availability to students, and serving on committees. Among other benefits to the community college, reformers believed that that greater connection to the institution would maintain greater consistency in the application of course standards and improve course outcomes for students. The present investigation tests if the employment status of the instructor has an effect on the outcomes of students.

A primary basis of this investigation is in public policies that are intended to protect educational opportunity and access to education at the community colleges in California. Educational opportunity and access are to be safeguarded by regulations that community colleges empirically demonstrate the reliability and validity of their placement tests and practices. In addition to California state regulations, there are also state and federal court decisions that reinforce the need for predictive validity measures of academic placement tests and practices in the open-access college.

In determining the predictive validity of instruments used for the purpose of grouping people by ability, institutions must demonstrate that individuals have a demonstrably greater likelihood of success above a certain "cutting" score (minimum score on a test) than individuals falling below this certain score. In some cases, opportunity for courses is allocated according to an inferred level of ability or aptitude as determined by completion of a prerequisite skill level in English or mathematics deemed necessary for success. Faculty and administrative leaders assumed that students achieving this prerequisite skill level, whether through prior coursework or a minimum score on a standardized placement test, have a greater likelihood of success than students without the

required pre-requisite. The present investigation study compares the outcomes of students eligible for a course, with the outcomes of students deemed ineligible for a course.

Predictive validity studies are often hampered by the pre-sorting of students into levels. Pre-sorting or screening students truncates the range of the student test and biographical characteristics. Truncation of range of the independent variables (such as test scores) can be found in community colleges where a testing and placement program exists. Therefore predictive studies are often severely constrained (Sheldon, 1970; Cronbach, 1971, State Chancellor's Office, 1992). Students with lower placement levels are often prevented from enrolling in higher level courses. A truncated distribution often results in lowered correlation coefficients between predictor and criterion (Anastasi, 1968).. The present investigation takes advantage of a time period during which mandatory placement rules were temporarily relaxed and students could enroll in courses they would otherwise have been ineligible to attempt. The relaxation of placement rules helped to reduce, but not eliminate, the problem of pre-sorting of students.

In the present investigation, particular attention is devoted to the assessment of standardized test score validity for predicting success when other data about instructors and students commonly available in a community college setting are used. The use of multiple indicators about individual aptitude for success or retention in college courses is made operational in this investigation. Integrating student biographical data into a model to improve course placement and retention is thus a major goal of this study.

This investigation affords the unique opportunity to compare the performance of students considered ineligible for a course with students considered eligible. One part of the analysis is given to the performance of those rejected for a course based on an inferred academic deficiency, but allowed to attempt the course. The success rates of the ineligible compared to the eligible can thus be compared. Comparing success rates for the two

groups of students is intended to inform the debate on the merit and application of mandatory placement rules when other potentially relevant information about the student is not employed in the eligibility decision.

The approach to this investigation was strongly guided by theory. Measurement theory was used to design a study that recognized the difficulty of predictive validity when the criterion variable is a measure based on individual judgment, when student characteristics differ, and when the characteristics of the instructor differ. The research questions and hypotheses for this study were informed by the constructs from Point-to-Point theory and Aptitude-Treatment Interaction theory.

Background

Although the debates surrounding testing, access, and misclassification are not new, the debates appear to be born again in every generation. Much has been written over the years to suggest that our assumptions about a person's ability to succeed in some future endeavor at some point in time is an imprecise science at best. This is particularly true when the criterion is unreliable (Cronbach, 1990; Cronbach and Gleser, 1965; Popham, 1990; Jensen, 1993; National Commission on Testing and Public Policy, 1991; Sticht, Armstrong, Hickey, and Caylor, 1987). However, during the late 1980's, educational reforms legislated for the two-year colleges in California not only mandated testing for placement purposes, but also mandated the minimum predictive validity coefficients of placement tests with performance outcomes in college courses (California Community Colleges, 1992).

Several statewide reforms affecting higher education were passed by the California state legislature during the 1980's. Among the reforms affecting the community colleges, the Matriculation Act (Assembly Bill 3, Chapter 1467: Seymour-Campbell Matriculation

Act) is of particular interest to this study. The Matriculation Act mandated that California community colleges design and implement a Matriculation program to improve student outcomes. Outcomes improvement would occur by more clearly delineating the responsibilities of students as participants in the educational process. The colleges were also mandated to provide enhanced support services to help students meet those responsibilities (Henrikson, 1995; Alkin and Freeman, 1992).

One of the most controversial components of the Matriculation Act was the component that mandated assessment and placement (Fields, 1988). As implemented in most colleges, the testing and placement component relies primarily on the use of standardized tests for student placement and tracking into different levels of English and mathematics.

The implementation of Matriculation reform in the California community colleges helped to fuel rapid growth in the use of nationally standardized and locally developed tests for placing students into different levels of the college curriculum. A recent survey of 101 California Community Colleges showed that 92 colleges were using a standardized placement test, or a modified version of a standardized test. Nearly all respondents had set cutting scores to place students according to abilities in reading, writing, and mathematics (Armstrong, 1997). The scores from standardized tests are also used to determine student readiness for other courses in the curriculum such as the sciences, foreign languages, and vocational programs.

Student course eligibility on the basis of an inferred skill level from a test score or completion of a prior course has been receiving much attention lately in the community colleges, particularly with regard to students thought to need developmental and remedial courses (Fonte, 1997). Within the Matriculation reform was the authority for colleges to require course prerequisites for entry into selected college courses. Determining eligibility

for a course based on a test score based inference or prior course completion had the effect of restricting entry into college courses for many students. Mandatory communication (English or reading) or computation (mathematics) prerequisites can either be based on a score (such as a test score), completion of a class, or an inference about student ability based on individual judgment by an instructor.

Growth in standardized placement testing combined with a renewed emphasis on the prescriptive use of scores to allocate opportunity led to increasing anxiety by various groups both in and out of education. Apprehension over the growing use of standardized tests reflected decades-long distrust and worry over the potential misuse of testing and its effects on various student groupings (Gifford, 1989). Thus the notion of test anxiety did not appear to apply just to students preparing to sit for an exam. A collective test anxiety could be found among various social, advocacy, and testing watchdog groups as to how standardized placement tests would be used, interpreted, and applied in educational settings. This collective test anxiety soon made its way into court.

Not long after the passage of the Matriculation Act activist groups took legal action. In 1988 the Mexican-American Legal Defense Fund (MALDEF) alleged that several community colleges were using scores on placement tests to force students to take non-credit, remedial courses before they could enroll in courses carrying college credits that were transferable to four-year colleges and universities. The suit, which included 19 other colleges charged that these colleges violated the Matriculation Act and due-process guarantees of the California constitution by refusing to allow some students to enroll in college-level courses for credit. (Fields, 1988).

Advocacy groups opposed to mandatory placement testing also cited federal civil rights laws on the use of tests to screen and sort students. Federal statutes regarding the use of testing was codified in Section 9 of the 1991 Federal Civil Rights Act that focuses

specifically on test scores. Advocacy groups such as MALDEF argued that students from historically under-represented groups were unfairly affected by such practices. That is, there was a "disparate impact" on certain minority student groupings as a result of mass testing (Kaplan and Saccuzzo, 1982).

As with tracking and ability grouping debates in the 1960's and 1970's, debates over the use of tests to track students into levels of the curriculum took on racial and ethnic overtones as well. The colleges were forced to respond to challenges from social groups and advocates in the community concerning the use of tests for tracking and sorting minority students. The activist groups were responding in part to beliefs that standardized placement tests represented another method for systematically excluding and channeling the aspirations of culturally disadvantaged students. (O' Connor, 1989)

Critics of the use of testing were also joined by some academics and policymakers who view the community college as an institution that maintains social stratification and solidifies class differences. Critical views of testing and placement were compatible with beliefs that the community colleges had not fulfilled their economic and social obligation of uplifting entire groups of people. From the perspective of testing critics, standardized testing, sorting, ability grouping, and realistic goal setting for students were the mechanisms used to throttle student goals for transfer and further limit higher education opportunities (Brint and Karabel, 1985; Zwerling, 1986)

The MALDEF lawsuit cited above resulted in a settlement three years later. (Los Angeles Times, 1991). Partly as a result of that settlement the State Chancellor's Office for the California Community Colleges (SCOCCC) agreed to force every college using assessment tests for placement to gather and report content- and criterion-related validity evidence for each test they were using (Cage, 1991). If colleges were using locally developed tests, then colleges were required to report additional technical data such as test

item bias, reliability, and other information consistent with the Standards for Educational and Psychological Testing published by the American Educational Research Association, American Psychological Association and the National Council for Measurement in Education (American Psychological Association, 1985). Thus the state was compelled to monitor closely and approve the use of standardized placement testing in the two year colleges. SCOCCC regulations included a state-prescribed correlation coefficient between test scores and a criterion, and local college studies to report on the impact and the potential tracking effects introduced by the use of standardized placement tests.

Historically, the debate over the merits and problems of ability grouping and tracking had previously focused on the primary and secondary grades. However with the implementation and use of mandatory placement systems in the community colleges, the testing versus tracking debate became pertinent to the open access community college.

To provide guidance to the community colleges in the test validation process, the SCOCCC published a document in August 1992 entitled, *Standards, Policies, and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges* (California Community Colleges, 1992). This document sets forth criteria for colleges and districts using assessment tests for placement purposes. All colleges must provide evidence pertaining to the following standards with regard to tests:

Minimize or eliminate cultural bias,

Monitor disproportionate impact on particular groups of students by age, ethnicity, sex, or disability grouping,

Establish local norms,

Establish the local reliability and validity of placement tests,

Validity of "cut-scores" used for placement

The use of multiple measures in course placement

The predictive validity of the placement test to the outcome measure.

(i.e., demonstrate a validity coefficient of at least .35)

The state required that evidence must be gathered to address each of the above regulations for review by visiting technical assistance or state audit teams. For example, institutions that use an unmodified test for placement must demonstrate that students scoring above the cut score have a "relatively greater expectancy of success" (obtaining a "C' grade or higher) than students scoring below the score or score range (ibid. p. 21).

Similar polices exist for documenting the criterion-related validity of prerequisite courses or skill levels in the community colleges. Colleges are to provide evidence to describe and empirically evaluate the effects of requiring a prerequisite for enrollment in a course. The evidence demanded here for prerequisite validation is similar to that required for test validation: Colleges must show compelling evidence that eligible students have a far greater likelihood of success than ineligible students (Alkin and Freeman, 1992).

To respond to the civil rights questions brought up by testing critics required predictive validity studies and evidence on the impact of testing on certain student groupings. Interestingly, the observations of Cohen more than ten years earlier (1975), on how the community colleges would face increasing pressure for access brought to bear by community and political groups appeared to be coming to fruition in the late 1980's.

Federal law also requires institutions to gather and report the predictive validity of placement practices (Petrillo, 1971; Reutter, 1975). Federal court decisions in the 1970's required that institutions validate the use of assessment tests in the placement, advancement, and hiring of individuals (Linn, 1972). Federal law governs the use of tests or other score-based inferences for screening students and the relation of test scores to a placement, hiring or promotional decision (Education Commission of the States, 1980). As in the case of the California state laws, the federal statutes and court decisions were

intended to mitigate the perceived negative effects of mandatory placement testing and score-based inferences on the opportunities and access of students, job applicants, and other individuals.

Laws and regulations mandating predictive validity evidence was a direct challenge to common practice in the community colleges. No longer was professional judgment of course rigor and content review sufficient for educators to justify minimum cutting scores for access to courses or programs. Increasingly, state and community demands for accountability and objective, verifiable, evidence began competing with professional judgment for decisions regarding student eligibility and potential for success in college courses. Thus, at the same time as community colleges increase their reliance on testing and prerequisite courses as screening and sorting devices there are a mounting set of legal and ethical mandates for the colleges to document and report criterion-related validity evidence.

To comply with federal civil rights laws and state regulations, the California State Chancellor's office also requires that colleges use other measures in the placement decision. The purpose of the use of companion measures is to ensure that students are not placed on the basis of a single test score. Although the value of the use of a scale that integrated student biographical data with test scores has been discussed by measurement scholars for several years (Cronbach, 1990; Anastasi, 1968) and advocated by some professional organizations (American Psychological Association, 1985), many community colleges have been slow to identify or adopt companion measures for placing students (Alkin and Freeman, 1992). One goal of the present study is to integrate student biographical characteristics and test scores in a model to improve the prediction of success or retention in college courses.

Statement of the Problem

Standardized tests serve a wide range of purposes. Scores from standardized tests are used to measure program effectiveness (such as high school preparation or remedial courses), diagnose individual student needs for remediation, restrict entry to courses or programs, and certify that students have mastered certain skills.

Sticht (1990) suggested that the competing uses of the same type of test instrument may be a misuse of testing. Often the skills measured on standardized tests are not matched with objectives in the school curriculum. Another problem originates in the use of tests as diagnostic devices. It has been noted that the sampling of knowledge from standardized assessment tests is often insufficient for diagnosing student needs for remedial instruction (Jackson, 1990; Sheldon, 1970). Also, the use of tests to determine eligibility for courses or programs often ignores other potentially relevant information about the characteristics and traits of the student that may more accurately predict their chances for success or retention in a course.

Evidence of criterion-related validity is discerned by comparing test scores, or predictions made from them, with an external variable (criterion) that provides a direct measure of the characteristic or behavior in question. The simplest, most logical, and often the only method for community colleges to gather evidence for the predictive validity of tests is to compare placement test scores with another measure that provides evidence of academic success--final course grade or retention.

Although course grade or retention may be considered the most logical and administratively convenient criterion variables, they may not be the most reliable measures. Final grade or retention in the course may not meet the necessary requirements of a criterion variable as specified by measurement theory. Instability in prediction is more difficult when the criterion variable is a measure based on individual judgment. The uncertain nature of

commonly available criterion variables to measure course success combined with great diversity in student characteristics makes predictive validity of assessment instruments problematic. Other confounds to prediction such as test measurement error, varying instructional techniques and instructor characteristics, exacerbates the problem of predicting student performance in college.

In addition to studying the relation of test scores and ability levels to course outcomes the present study focuses on other student biographical variables that could be employed when institutions attempt to select, sort, and classify students to improve their chances for success. The inclusion and weighting of student demographic, situational, and dispositional variables in the placement decision is a problem facing many community college placement programs.

Testing proponents have suggested that student retention during a semester will be enhanced through the use of a placement testing system (Mount San Antonio College, 1989). The primary argument for this point of view is based on the notion of academic self-concept. This concept suggests that students placed in a group where they have roughly equal abilities with peers will given them enhanced confidence. Advocates argued that students, particularly those considered under-prepared or low achieving, would be less likely to drop out if they were placed in courses that were not too challenging for them (Kingan and Alfred, 1993).

Placement testing and grouping by ability level was supposed to improve the retention of students, particularly students considered academically underprepared. Underprepared students grouped together would develop more realistic expectations of their own readiness to do college level work (Manning, 1991). Testing advocates also suggested that less prepared students placed in the same classroom with highly prepared students tend to leave prematurely because they compare themselves to the better prepared

students. This comparison lowers their academic self-confidence. Lowered self-esteem leads to an increased tendency to drop out. Thus an argument advanced in favor of grouping students by ability is that classifying and sorting students would narrow the gap between academic self-concept and achievement. To test the value of placement testing on the retention of lower achieving students, the relation of test scores to retention is analyzed by level in the curriculum. Courses from transfer level to non-credit basic skills are included in the analysis. Including non-credit courses in the study allows for the testing of the relationship between test scores and retention for the less-prepared students.

However, as with predicting grades from placement tests, using scores from placement tests as predictors of retention is problematic. Although attention has been given in prior studies to the relationship between admissions test scores and college outcomes such as freshman grade point average and retention, research has been largely restricted to four year colleges and universities (Pascarella and Terenzini, 1991). Comparatively little research has focused on this phenomenon in the community colleges. Predicting retention from placement scores is seldom mentioned or tested for community college students. The present investigation will help to fill a gap in the research literature by examining the relationship between standardized test scores and retention.

It has been noted in the literature that predictive validity studies are sometimes muddled or inconclusive because of unreliable or inadequate criterion data (Kaplan and Saccuzzo, 1982; Popham, 1990; Cronbach, 1971). The validity of a placement system that sorts and classifies thousands of students is often reliant on individual judgment provided by instructors, raters, or other evaluators of the student's performance. The use of individual judgment or rating to determine the predictive validity of a score or measure often introduces error into the calculation of the predictive validity coefficients (correlation

between test score and final grade or retention). Controlling for this error is desirable to understand and improve the effects of placement decisions on students.

To control for variance in the criterion variables of final grade and retention, the present investigation includes a potentially rich source of variation, the instructor, as part of the model predicting course grades and student retention. Including the instructor in the course performance outcomes models will quantify the amount of variance in final grade or retention that can be attributed to instructor differences. The analyses on the instability of the criterion used to judge placement programs should inform test validation policy and practices.

Until the present study, there has been little evidence gathered that indicates how much error in the predictive validity of a test or other measure may be contributed by the individual judgments made by instructors, employers, or trainers. In discussing the usefulness of grades or marks as criterion data for evaluating placement decisions using tests, Cronbach (1971) makes the following point: "If teachers use different bases for judgment and some are more generous than others, throwing grades from several algebra teachers into a single distribution merely piles one source of error on another" (p. 491). When a test fails to accurately predict student course performance, it is difficult to determine whether this observation derives from the test, the grading process, or other factors such as motivation, experience, and study habits, that also affect course performance.

Purpose of the Study

The purpose of this study is to examine and analyze test score-based inferences commonly used to predict success in the community college environment. Standardized test scores are commonly used for this purpose, but it is necessary to assess the validity of

this practice. Another common practice is to predict success based on student completion of a course considered a prerequisite for another course, thus restricting access to a group considered prepared by completing the required prerequisite. A required prerequisite course is often a pre-collegiate English or mathematics course or a cutting score on an English or mathematics placement test. A prerequisite English and mathematics skill is needed so that students are able to read, write, and calculate at a level thought sufficient to negotiate course readings and assignments. The practice of classifying students as eligible or non-eligible for a course based on an inferred level of English or mathematics ability is increasing in many community colleges (Berger, 1997). Therefore the predictive validity of the classification of students as eligible and non-eligible based on their completion of prerequisites also merits examination.

This investigation applies theory to analyze testing, measurement, and predictive validity in a community college environment. Several theories pertinent to this study topic were examined. These included Classical Test Theory, Decision Theory (Cronbach and Gleser, 1965); Aptitude-, or Trait-Treatment Interaction Theory (Willingham, 1974;) Point-to-Point Theory (Asher and Sciarrino, 1974;) and Identical Elements Theory (Thorndike, 1904). Review of theory contributed to the formulation of the research questions, hypotheses, and course outcomes model used to explain variance in college course outcomes of final grade and retention.

Of the several theories reviewed, Point-to-Point theory and Aptitude-Treatment-Interaction theory emerged as the most useful to this investigation. Point-to-Point theory was found to be the most instructive in the formulation of research hypotheses and positing relationships between the independent variables of student test scores, demographic, dispositional, situational characteristics and the dependent variables of final grade and retention. Aptitude-Treatment Interaction (ATI) theory provided a solid theoretical rationale

for analyzing instructor effects on the criterion variables of retention and final grade. Using constructs derived from Point to Point and ATI theories, a framework for the investigation was constructed. The theoretical framework enabled the present investigation to take a closer look at the validity and usefulness of placement tests, and to posit relationships among student characteristics and instructor effects in explaining variance in course outcomes in the community colleges.

Relating testing and completion of prerequisite courses to retention outcomes has been given attention in the literature (Wall, 1996.). With respect to student retention, much of the literature identifies various causes of student retention, most having to do with student involvement, engagement, integration with the institution, and development (Astin, 1975; Bean, 1980 and 1982; Lee and Frank, 1989; Pascarella 1982; Rogers, 1990; Tinto, 1993; Spady, 1971). Much of the student retention literature is focused on students attending four-year colleges and universities of varying size and selectivity. Few of the retention studies reviewed seemed applicable to studying retention or performance in the two year, open-access college.

Constructs of student characteristics were identified and grouped using a model described by Cross, (1981). Cross identified three primary factors related to course outcomes among adult students which appear to be particularly applicable to the present investigation. These factors were situational variables (time, employment, child care, transport), institutional (time needed to complete program, inconvenient class scheduling, strict attendance requirements), and dispositional (self-efficacy, past experiences in school, support from friends and family). The present investigation uses questionnaire and institutional records data to measure aspects of Cross's model for student retention and final grade. Of particular interest were student dispositional and situational variables that Cross suggested caused many students to drop out of or perform poorly in school.

Importance of this Study

In light of the growth and debates over the use of standardized placement testing in the community colleges, the present investigation is timely. Recently Kirst (1997) observed that in California higher education the lack of standardization in admissions and placement testing combined with a failure to behaviorally link tests with state curriculum guidelines creates problems for students and schools. He suggests that the debate over standardized tests for admissions and placement is at the core of a general debate over what knowledge is most worth possessing. Kirst argues for better alignment of placement tests used in the three segments of higher education in California (University of California, California State University, and California Community Colleges) with the K-12 assessment tests and state curriculum guidelines. Test and curricular alignment may help to improve the correspondence between what is taught in school and what is measured on tests.

Demands for test accountability and validation evidence to justify allocation of opportunity are increasing in many public arenas. These arenas include education, employment, occupational training, access to military services, and professional associations. For example, according to contemporary federal standards for equal employment opportunity, tasks used to test potential new employees must demonstrate relevance to actual job performance. Thompson and Thompson (1982) reviewed 26 federal court decisions where the validity of tests used to screen employees had been challenged. In their review, the authors found that the job-related test must focus on tasks, should be in a written form, and must include several different data sources with large samples. Increasingly, judgments rendered under current case law are requiring empirical evidence that a score-based inference used to screen applicants is valid for predicting proficiency in a job or in training programs (Kaplan and Sacuzzo, 1982).

Recently, there has been a strong focus on both the state and federal level on the uses and impact of testing as a tool to allocate access and predict future behavior. A review of documents from the California State Education Code (1990) and related statutes stress the importance of the use of testing as a predictor of future performance. Federal statutes and court decisions indicate that education and training institutions will be called upon to demonstrate that admissions and course entry standards are justified and essential for individuals to succeed. Court decisions increasingly require that test developers and users must strive to match the predictor to the eventual criterion to use the tests as screens for employment. In the community colleges however, such a practice is not routinely followed (Cohen, 1987).

It has been suggested recently that the movement toward race-neutral admissions policies and school effectiveness measures that both rely heavily on standardized testing will soon collide with federal civil rights laws (Brownstein, 1997). The use of supposedly race-neutral measures such as test scores can violate federal civil rights law if opportunity is differentially allocated to minorities or women. Institutions using test scores to admit or place students into programs may have to justify the setting of minimum test scores as a condition for admission to a program or course of study. This may occur if, as recent trends suggest, certain groups do not perform as well on such tests and evidence of a disparate impact on admissions can be found.

The importance of this study is further underscored in light of an increase in state mandates, court decisions, and public policy implications of standardized tests to allocate opportunity and access. Providing predictive validity data is no longer an option for many community colleges and other institutions, particularly in California. Calls for test accountability will continue to resonate as placement testing used to screen and sort individuals continues to expand.

The issues of placement testing and mandatory placement are not confined to California. In other states, use of tests as screening and certification devices is also growing in the community colleges. For example, the Texas state legislature recently mandated the use of a placement test for use in all community colleges in the state to determine which students need remediation (Rucker, 1997). Students not achieving a sufficient score on the Texas Academic Skills Placement (TASP) are required to enroll in remedial courses until they are able to pass the test. If they do not enroll in the remedial English or math class, they are not allowed to attend classes at the college even if the student passed the remedial course with an 'A' grade. Texas policymakers appear to be wary of the use of grades as a measure of student learning. Also the Texas remedial policy tends to suggest that there are strong societal beliefs in the validity of testing both as an assessment of school effectiveness, and as a diagnostic tool to measure presumed student academic deficiencies. The present investigation will help to inform this debate.

Other states are considering the adoption of minimum test scores for students in college remedial programs to enter college level lower division courses. Citing the high costs of remediation, recently the Florida legislature has voted to raise the minimum cutting scores for entry into lower division courses in the Florida community colleges (McCabe, 1995). The student may have achieved a passing grade in the course, but an insufficient score prevents the student from enrolling in college level courses. One goal of the present investigation is to help inform the debate over the use of standardized tests such as those in Texas and Florida as indicators of individual aptitude and predictors of future success in college.

External demands for predictive validity evidence comes not only from government agencies and courts, they also come from regional Accrediting associations. For example, recently issued guidelines for Accreditation (1996) indicate that community colleges using

placement tests as a sorting device for different levels of education or training provide evidence that placement tests or rules are fair, accurate, and valid (Western Association of Schools and Colleges, 1996). In Accreditation Standard 5, one indicator of good educational practice is that assessment tests and placement practices are valid, fair, and equitable.

Test validation policy is increasingly coming into conflict with faculty prerogative over curriculum and courses. Reform legislation passed in the late 1980's gave community college faculty primary responsibility for curriculum policy in the community colleges (Assembly Bill, 1725, 1986). The California state legislature gave faculty the authority to mandate that students complete courses in a specified sequence. In addition, entry into a course could be restricted unless students demonstrate competence according to scores achieved on placement tests. Many academic departments chose to restrict student enrollment in certain courses or programs to students who demonstrated a certain skill level or had completed a prerequisite course. However, state testing and predictive validity guidelines has moved the decision on restricting access from the qualitative to the empirical. Community colleges may restrict access to courses, but only if they can show that students lacking the prerequisite are highly unlikely to succeed. State interest in the predictive validity of testing is moving test validation closer to the community college classroom door.

The problem of misclassification is pivotal to the public debate over access. What happens to those individuals rejected or declared ineligible for a course, program of study, or employment? If they were given the opportunity to enroll or challenge a course how does their performance compare with eligible students? The present investigation offers the opportunity to answer these misclassification questions by comparing the course outcomes between eligible and non-eligible students. During the selected timeframe for this investigation, many potentially ineligible students were allowed to attempt a course and

earn a grade. Research designs that attempt to validate course entry requirements are often confounded by a lack of ineligible students allowed to attempt a course. The unique conditions afforded by the temporary relaxation of entrance requirements in the colleges provided the opportunity to gather evidence for the effect of course eligibility on success.

Apart from the need to gather and document evidence of fairness, accuracy, and validity in the use of cutting scores or prerequisites for legal reasons, it is also the ethical course to take. The integrity of our educational institutions rests to some extent upon the placement or access decisions mandated for our students. If there is sound empirical evidence that a particular cut-score or skill level is essential to student success in a particular course or program, then enforcing the prerequisite or skill level is a sound educational decision. If however, the evidence does not suggest that students without the mandated prerequisite or cut-score are highly unlikely to succeed, then blocking student enrollment and compelling them to enroll in remedial courses is unfair and arbitrary.

The present investigation is intended to provide guidance to other researchers and evaluators seeking to gather and analyze criterion-related validity evidence for the use of placement tests or prerequisites. Findings from the investigation may be particularly helpful when the criterion for validation is final grade or retention in the course. Thus the present study will also help to fill a gap in the research literature on test validation and student retention.

This investigation may also stimulate thinking on the poorly understood and erratically defined concept of the underprepared student. It seems that definitions of underprepared depend much on who is doing the defining and under what conditions. This situation can be found in many institutions of higher education. When enrollment is slack, some institutions adjust entry standards or create special programs to admit students who otherwise would be rejected from entering the university. When there is a surfeit of well

qualified students, entrance standards are often raised and many begin to question if the remedial role is appropriate for the university.

The term "academic standards" as it is used by many educators merits attention. One intent of the present investigation is focus more closely on the concept of academic standards as the term is commonly used in higher education. In the context of higher education, academic standards are often viewed as something imposed on the college or classroom by the students, rather than imposed on the students by the institution. It appears that when educators speak of declining academic standards, they are often referring to the perceived level of educational preparation and personal traits of students. For example, the authors of "A Nation at Risk" (National Commission on Excellence in Education, 1983) recommended that colleges and universities raise admissions criteria to reverse a perceived decline in academic standards. The "rising tide of mediocrity" could be slowed or reversed by a breakwater fashioned of higher admissions criteria and test scores. It appears that there are widely held assumptions that academic standards are reflective of the selectivity of the institution, and that open-access institutions have lower academic standards because of the generally lower entering abilities of students. This study will help to challenge the assumption that academic standards can be raised by increasing test scores and requirements for college access.

Research Questions

The research questions and hypotheses derived from them are guided by an analysis of current social and historical forces, measurement theory, and recently enacted educational policy and reforms in the California Community Colleges. The present investigation centers on three primary research questions.

1. Are placement tests highly predictive of course performance outcomes such as course grades or retention in the class?
2. How do student characteristics affect the prediction of performance outcomes?
3. How do teacher characteristics affect the prediction of performance outcomes?

The research questions led to the development of four primary hypotheses. The null and alternate hypotheses guiding the present investigation are presented below. Following the hypotheses is a discussion of the assumptions and rationale of each of the alternate hypotheses.

Hypotheses

The null and alternative hypotheses for this investigation are:

Null Hypothesis 1:

There is no relationship between course performance outcomes as measured by course grade or retention and scores on standardized placement tests in community college English and mathematics courses.

Alternative Hypothesis 1:

There is a relationship between course performance outcomes as measured by final grade or retention and scores on standardized placement tests in community college English and mathematics courses .

Alternative Hypothesis 1a:

The predictive validity coefficient as measured by final course grade or retention will fall below the threshold level of between .30 and .40 recommended by Cronbach

(1975, p. 429), and below the .35 correlation coefficient mandated by California state education code for test validation.

Alternative Hypothesis 1b:

Scores on successive administrations of the placement test will show a greater point-to-point relationship (i.e., a higher predictive validity coefficient) than the coefficient between test score and final grade.

Null Hypothesis 2:

There is no relationship between course eligibility status (i.e., students are either eligible or not eligible to enroll in the course) and success in selected community college courses.

Alternative Hypothesis 2:

There is a relationship between course eligibility status and success in selected community college courses.

Alternative Hypothesis 2a:

The effect of eligibility status on success in selected courses will be small to modest .

Null Hypothesis 3:

There is no relationship between student background biographical characteristics and course performance outcomes as measured by course grade and retention after controlling for test score.

Alternative Hypothesis 3:

There is a relationship between student background and biographical characteristics and course performance outcomes as measured by course grade and retention after controlling for test score.

Null Hypothesis 4:

There is no relationship between instructor characteristics and final grades and course retention after controlling for student characteristics and test scores.

Alternative Hypothesis 4:

The inclusion of the instructor as a source of variance will contribute significantly ($p < .05$) to a model predicting course grade or retention after controlling for student characteristics and test score.

Alternative Hypothesis 4a:

Teacher characteristics will account for more variance in the course outcomes of final grade or retention than the simple correlation of test scores or retention.

Alternative Hypothesis 4b:

There is a relationship between instructor status (full- or part-time employment) and student final grades and course retention after controlling for student characteristics and test scores.

Several of the hypotheses described above focus on the central question of the predictive validity of placement tests with respect to student course outcomes. As suggested by Alternative Hypothesis 1, there are theoretical reasons to expect that the observed coefficients will be low. Theories concerning the relation between predictor and criterion such as Point-to-Point theory, Common Method Variance, and Identical Elements suggest that the evidence for criterion-related validity will be modest at best. This is due to the lack of common elements between the predictor (test score) and the criterion (grade or retention). Lack of commonality will lower the coefficient.

Alternate Hypothesis 1a focuses on the magnitude of the predictive validity coefficient between test score and final course grade. Cronbach (1975, p. 429) suggested that the minimum threshold of a predictive validity coefficient ought to be at least .30 for

the test to be of much value. A similar observation was made by Anastasi (1968). The state of California appears to have reached much the same conclusion as evidenced by published regulations mandating a .35 correlation coefficient between predictor and criterion. The .35 coefficient will be used to determine if the placement tests used in this investigation meet Cronbach's recommendation and state guidelines as evidence for Alternate Hypothesis 1a.

Alternate Hypothesis 1b focuses on the point-to-point correspondence of predictor and criterion. Specifically, this phase of the investigation will test if scores on two versions of a placement test, one administered at the start of one semester and another at the beginning of the next semester, have a higher correlation coefficient than the coefficient between test scores and final grade. Point-to-Point theory suggests that the relationship between test scores and course performance outcomes of final grade and retention will be modest due to a paucity of common elements. However this same theory would also posit that the correlation between scores achieved on successive administrations of a placement test should be higher than that observed between test score and final grade. The correlation coefficients are compared with the coefficients observed between test scores and final grade as evidence to accept Alternate Hypothesis 1b.

Hypothesis 1 also includes retention as a dependent variable. Prior research has suggested the decision to remain enrolled in a class or in college has several facets and is influenced by many factors often not measured or gathered by standardized tests (Astin, 1975; Tinto, 1993; Pascarella and Terenzini, 1991; Pascarella, 1982). A major argument in favor of using tests to group students is that retention rates will improve if students receive proper and accurate placement in a course that challenges but does not overwhelm (Kingan and Alfred 1993). Institutional application of the assumptions of decision theory suggests that student propensity to drop out or leave college prematurely is diminished when they can successfully negotiate course assignments and keep up with assigned readings

(Willingham, 1974). This argument has been articulated forcefully and persuasively by those who militate for ability grouping in the primary and secondary schools. A summary of these arguments can be found in Oakes (1985) and Tyler (1974).

The problem of student misclassification is the focus of Alternate Hypothesis 2. State policies mandate that community colleges provide evidence that students without the required prerequisite course or test score are "highly unlikely to succeed." (California State Matriculation Regulations, 1992). Alternative Hypothesis Two will determine the success rates of students who would have been excluded if not for the temporary relaxation of the mandatory placement system. It is possible that many students could successfully complete a course through dint of other dispositional or situational factors not measured or taken into account by the standardized placement tests.

In the colleges where this study is being conducted, local policy mandates that in order for a course prerequisite to be considered valid, students with the prerequisite or mandated ability level to enter selected courses must succeed at twice the rate of students without the mandated ability level or prerequisite. In addition, local policy requires that no more than 33% of the students without the prerequisite are successful in the course in order for the pre-requisite to be made mandatory . These standards are meant to lessen the number of misclassified students and false positives. These standards are also consistent with other community colleges in California (Chabot College, 1996; Isonio, 1996).

Hypothesis 3 focuses on student level variables in addition to test scores that can help to improve the prediction of student course outcomes. Does demographic, dispositional, and situational information about the student improve the fairness and accuracy of college placement practices?

Hypothesis 3 is also based on public policy mandates in the area of testing. For example, one outcome of the legal challenges to testing in the community colleges was a

renewed emphasis on the use of multiple criteria for use in the placement decision. Although recently mandated under state statutes, the value of multiple indicators for improving placement was noted three decades ago by Cronbach and Gleser (1965).

To assure testing critics and the courts that students would not be placed solely on the basis of a single score, California state officials intended that other academic or biographical data would be used to yield a composite picture of the matriculating student (California Community Colleges, 1992). Companion measures used with test scores for placement might include prior academic preparation, high school grades, number of hours worked, and information obtained through a review of student transcripts. State officials intended that this information, in conjunction with English and mathematics test scores, would be used as evidence of aptitude for success at a particular level in the curriculum. Test scores were eventually intended to be part of an overall assessment of individual aptitude, not the sole criterion of ability to benefit from instruction at a certain level. However, several years after this regulation had been implemented, many colleges still had not properly complied with the multiple predictor regulations. (Alkin and Freeman, 1992).

As suggested by Alternative Hypothesis Three, it is expected that including student biographical information with test scores will improve the prediction of student course performance outcomes. Theories reviewed for this study suggest that model integrating student biographical information with test scores will raise the observed coefficient above that of the proportion of variance explained by test score alone.

To test Alternate Hypothesis 3, the present investigation developed a course performance outcomes model to improve the quality of the placement recommendation. Using student level constructs described by Cross (1981), the present investigation developed a model that identifies the situational and dispositional characteristics of study

participants. Student characteristics are grouped according to Cross's schema and used in the model for explaining variance in student course outcomes.

Alternative Hypothesis 3 also posits a significant relation between the dependent variable of student retention, and the independent variables of student biographical data and test score. There are theoretical reasons to expect the prediction of retention to improve using multiple measures or variables in the course performance outcome model. In a review of theories related to student retention Bean (1982) noted that many student retention models contained several variables pertaining to student demographics, academic preparation, aspirations, commitment to college, intent to persist, and certain cognitive variables. A comparison of the Bean and Cross models suggest that both have much in common. For example the student aspiration and commitment variables identified by Bean appear similar to the dispositional variables identified by Cross. Many of the constructs used to define student traits and characteristics were available to the present study through a questionnaire administered at the time of placement testing.

Hypothesis 4 focuses on the stability of the criterion variable of course grade. Obtaining reliable and valid criterion related evidence for placement testing has been problematic. In educational settings where tests are used to classify and select students for placement, grades are often the criterion variable used to judge the validity of the placement decision. However what is overlooked in many studies is the identification of significant confounding variables in establishing criterion-related predictive validity. Unreliable criterion data to judge a placement system is particularly important if significant variation in course grade, rating, or retention practices outcome can be attributed to the individual assigning a grade (i.e. the instructor). Individual variation by raters introduces a potentially significant source of variance in the measurement of the actual effect of prerequisite knowledge or ability on student performance in the course. If high variation exists in the

criterion variable (grades or ratings), how does variation affect the predictive validity of placement tests or a model that includes test data in addition to other predictors? The present investigation answers this question by including the instructor in the model as a source of variance.

Summary

This chapter has presented an introduction to the present investigation. The current agitation over the use of testing in the community colleges reflects contrasting points of view of how society allocates opportunities and rewards. The testing debate has flared from the earliest uses of testing in Chinese civil service centuries ago (Miyazaki, 1981), to the present. The debate appears to be not over whether tests will be used, but how. Absent entry level placement testing, it is argued that institutions will have to invent new ways to select people. To borrow from a famous quote, placement testing is like democracy, it appears unattractive until you consider its alternatives. The present investigation intends to offer some alternatives to the current placement testing model by including other relevant student and instructor data that will improve the predictive validity and fairness of the testing, sorting, and placement process in the community colleges.

This dissertation is divided into six chapters. Chapter One introduces the study, provides background, purpose, statement of the problem, and presents the research questions and hypotheses. Chapter Two reviews the issue of placement testing in the context of the historical, social, and economic role of the community colleges in the US system of higher education. Chapter Two also examines the historical antecedents of testing and the general beliefs about the nature and distribution of intelligence and ability. This discussion will help to provide a historical foundation for testing and understanding currently held assumptions about knowledge and aptitude for college level work.

Perspectives on that role from competing theories in sociology - conflict and functionalist - are used to juxtapose contrasting points of view of the role of the open-access college in promoting equality and the uses of standardized placement testing. Prominent in this review in Chapter Two is the role of educational testing as a flashpoint in the debate over access and excellence in open-access colleges. Thus Chapter Two is intended to provide a context for this investigation into understanding the current upheaval over the use of placement testing and required skill levels for community college courses.

Chapter Three reviews relevant measurement and testing theories as they relate to predicting performance from scores and other inferences about individual ability. Chapter Three also reviews other predictive validity studies conducted in community college settings and discusses these findings in the context of the present investigation.

Chapter Four will present the methodology and logic of this investigation. Chapter Five will present the data analysis and the course performance outcomes model for explaining variance in course grade and retention. Chapter Six will present conclusions, recommendations, and implications for policy development and further investigation in the area of predictive validity in education and training institutions.

CHAPTER TWO

REVIEW OF HISTORICAL AND SOCIOLOGICAL FOUNDATIONS OF CURRENT DEBATES OVER TESTING IN THE COMMUNITY COLLEGE

"Testing of abilities has always been *intended* as an impartial way to perform a political function-that of determining who gets what." Lee J. Cronbach (1990)

Tests are increasingly used to provide an objective and common metric for grouping, classifying, and making inferences about the abilities of an individual to succeed in some future task or endeavor. The common metric provided by testing is desirable because of the increasing heterogeneity of the population and the need to rank people according to some consistent referent or scale. Tests provide a normative scale where an individual is compared against all others as a group to determine their relative position in a distribution. A common metric is desirable because previous methods or models used to make inferences about ability may no longer seen as reliable. For example, in the 1980's, the high school diploma began to lose credibility as an indicator of individual achievement and aptitude. (Morante, 1989; Roueche, Baker, and Roueche, 1984).

A standard score is also desirable when officials, recruiters, employers, teachers, or policymakers, need an objective method of sorting people that is not influenced by the views and biases imposed by a rater. As observed by an early proponent of testing, tests do not recognize rags nor riches nor position in social ranking (Cronbach, 1990). Rather tests are seen as a relatively simple, straightforward, method to document learning growth, evaluate program effectiveness, and assess individual achievement and ability. However in the view of some scholars and analysts, tests are increasingly used to allocate opportunity and reinforce existing rankings of social order and class (Owen, 1985).

Depending on one's point of view, the use of tests to derive inferences about individual ability to succeed in some future task either provides greater access and opportunities for success, or arbitrarily denies opportunity because of misclassification or inherent bias. Although not a new issue, the debate over whether testing serves as a gateway for student success or a gatekeeper that allocates opportunity according to the properties of the normal distribution continues to intensify. Testing has become a hotly debated topic that increasingly extends beyond the realm of the classroom, college campus, and worksite. Critics contend that testing has been used to inaccurately and unfairly relegate some to an undesirable treatment or prescription (such as a remedial reading course) while not identifying others who could profit from the same treatment. Supporters contend that testing has often been used to prevent improper assignment of individuals to unneeded treatments that may have occurred had they been rated by some human observer.

As the use of testing grows in traditionally open-access institutions such as the community colleges a conflict has arisen. Some believe that placement tests limit opportunity and unfairly track students. Others believe that tests are essential to maintain or reassert academic standards. Is there a fundamental conflict between testing and mandatory placement and the open-access mission? If there is conflict, what are the historical antecedents of this tension? What are the dynamics of this tension between the use of tests and mandatory prerequisite courses to screen and place students, and the open-access values and principles espoused by community colleges leaders? This chapter provides background and historical information to help illuminate the underlying reasons behind the often vociferous debate between testing proponents and critics in the community colleges.

The debate over testing also comports well with two competing perspectives on the community colleges from sociology. These two theories are functional theory and conflict theory. To functional theorists, testing, if properly done, reflects a rational approach to

identifying students of divergent abilities to improve their chances for success in school (Hughes and Nelson, 1991; Roueche and Archer, 1979). However, to conflict theorists, testing is another barrier to opportunity and replicates the artificial privilege accorded by social class distinctions (Collins, 1979; Owen, 1985). Conflict theorists point to the relationship between social class and test scores, suggesting that lower test scores are related to lower levels of cultural assimilation, not ability to succeed in post-secondary education.

Placed within the context of conflict and functional theories, standardized placement testing provides a fulcrum for these competing points of view in the debate over open-access versus raising standards in the community colleges. This chapter will include a section devoted to discussion of these two theories, and the importance of testing to each of them.

Historical Foundations of Testing in the Community Colleges

Cohen (1975) noted that primarily because of the open-access mission, local governance, and populist foundation of the community colleges, they are much less insulated from societal demands. This proximity to the community is important when considering how the colleges are affected by and respond to community and state intrusion into college operations and practices. Proximity refers to the isolation of a particular institution from community or state influence. The relative isolation of certain institutions lessens their proximity and hence reduces influence from community members, activist groups, and state legislators. In California for example, the University of California system has greater insulation from community demands because of constitutional limits on what the state legislature can prescribe for it. The appointed Regents of the University, although confirmed by legislative action, are, in effect, the sole governance body of the

university. Cohen suggests that because the community colleges have the closest proximity to the communities they serve, they will be among the first institutions to have to respond to pressing social and economic issues of the day, whether this is affirmative action, accountability, economic development, or immigration.

The contemporary debate in the community colleges over testing provides an example of the proximity concept. The historical proximity of the open-access colleges to the community helps to explain the imposition of state-imposed test validity regulations. For example, neither the California State University or the University of California are required to conduct validation studies of placement instruments, yet placement testing is used regularly to sort students into pre-collegiate English or mathematics courses in both university systems.

The Community Colleges: Access for All

As the premier open-access institutions in the US higher education system, community colleges have for the last several decades emphasized growth in student entry and recruitment. The evolution of the community colleges into comprehensive open-access institutions was a result of efforts to "...seek, recruit, enroll, and retain every possible student in the community" (Roueche, Baker, and Brownell, 1971, p. 11). Under this definition, the access mission was largely successful (Cohen, 1996; Deegan and Tillery, 1985; Breneman and Nelson, 1981; Knoell, 1980). In the years since 1960 community college enrollments have soared (National Center for Educational Statistics, 1996). Open-access policies combined with virtually unlimited growth in enrollment, development of new programs, course offerings, degrees and certificates, convenient locations, and liberal funding policies, all proved effective in creating unparalleled growth in the popularity of the community colleges (Deegan & Tillery , 1985).

Enrollment by groups traditionally under-represented in higher education such as women, minorities, and the disabled, has also grown tremendously (National Center for Educational Statistics, 1996). Low cost, few requirements for admission, flexible scheduling, and course offerings that were diverse in content and purpose brought in record numbers of students of diverse backgrounds (Cohen, and Brawer, 1996).

The ever-widening scope of the comprehensive mission was also promoted by funding formulas that rewarded high enrollments irrespective of course content or learning outcome (Breneman and Nelson, 1981). The debate was not over access versus quality, rather the general theme of the late 1960's and 1970's appears to have been how to accommodate all students who presented themselves to the college. There was a gradual but perceptible drift away from what Cohen (1980) called the "sequential curriculum" in which a community college education included a sequence of intended learnings in a pattern or progression that had some deliberate arrangement and order and that the intended learnings had outcomes that were predefined. Thus students could enroll conceivably in practically any course that interested them.

The Multiple Roles of the Community College

The comprehensive community college serves several important roles in the American educational system. These roles and functions have evolved and expanded greatly since their inception around the turn of the century. Bogue (1950) noted that in 1922 at their annual meeting, the American Association of Junior Colleges (AAJC) defined the junior college as "...an institution offering two years of instruction of strictly collegiate grade." However, three years later the definition had been significantly expanded. Besides the lower division emphasis, the AAJC also added that the junior college could also include different curriculums that were "suited to the larger and ever-changing civic, social,

religious, and vocational needs of the entire community in which the college is located." (p. xvii).

Since the time of Bogue's observations, many more changes have occurred in the role and purposes of the community college. In their description of the American community college, Cohen and Brawer (1996) describe an institution much expanded in breadth and complexity. They describe five general areas of the community college; career education, compensatory or remedial education, transfer and lower division education, general education, and community education and community outreach activities.

Cross (1985), described the multiple roles of the comprehensive community college as comprising both a vertical integration function and a horizontal integration function. The vertical function emphasizes articulation with high schools and transfer to senior level institutions. The horizontal function emphasizes community education that responds to the educational needs of a wide range of people and organizations.

Clark (1980) noted also the expanding mission of the community colleges, particularly in the western United States. Community colleges were continually redefining their boundaries where "...almost anyone, part- or full-time, can enroll in courses offering transfer credit..." (p. 20). He suggested that the redrawing of the educational boundaries by community college leaders may result in a more diffuse and confused role in higher education for the community colleges in the years ahead. The lack of clarity over the increasingly expanded community college mission was also echoed by Hendrick (1980).

The Social Role

Concomitant with the evolution of the comprehensive community college's education and training roles has been a growing emphasis on the social role. Cross (1986, p. 35) describes this function as the amelioration of past educational inequities and the

"preparation of students for citizenship, family life, cultural and esthetic appreciation, and lifelong learning." Cohen, (1987) also noted the importance of the social role to the community college mission. He notes that when the colleges are criticized, it is often for their failure to assist significantly in reducing economic and social inequity in the United States. Clark (1980) observed the strong effect of populist beliefs in the United States on how access to higher education is viewed. He noted that American interpretations of educational justice "equate equity with open doors" (p. 19). The open-door philosophy is an ingrained part of the social fabric in the United States, particularly in the California community colleges.

The social role and the rhetoric surrounding it is given voice in legislatures, community college catalogs, and pronouncements by community college officials and organizations. The community colleges' role in providing a bridge to economic betterment figures prominently in debates over access to college programs, and the outcomes and results of these programs. Increasingly, the social role has moved to the forefront of discussions of community college effectiveness and debates over the mission and priorities. It is perhaps the social role that causes the greatest conflict with the use of standardized testing in the community college.

The social role of the community colleges has its roots in Jeffersonian beliefs about equality of opportunity, personal economic betterment, and individual initiative. A cornerstone of the social role is the idea of building an aristocracy based on merit and achievement from a democracy of opportunity. In the view of an early proponent of mass testing in the United States, the closing of the physical frontier in the US made education the new frontier, where equality of access to opportunity could still be guaranteed. An objective metric provided by a standardized test score reflected both effort and ability. These were the keys to success in a democracy (Lemann, 1995). The community college is

seen by many as one of the few truly egalitarian institutions in the American education system. Therefore any attempt to limit access, sort students, restrict enrollment, or introduce explicit forms of tracking using test scores is viewed with suspicion and sometimes responded to with legal challenges.

The social and egalitarian function of the community colleges is a strong theme in the authorizing and subsequent reform legislation of the two-year colleges, the California State Master Plan for Higher Education, and current and historical descriptions of the community colleges (Cohen and Brawer, 1996). As stated by Cross (1986), a central task of the community college is "not to teach those who will be successful, but to make successful those who come." These same sentiments are echoed by national community college leaders (Wenrich, 1995), college mission statements and philosophies, and legislation. The egalitarian role of the community colleges is based on access to all individuals regardless of social position, academic history, aptitude, or achievement.

Education as an investment in individual human capital is deeply rooted in US cultural beliefs (DeYoung, 1989). The general reticence of the American public to support efforts for large scale income redistribution or transfer payments to the poor is reflected in the belief that access to low cost and accessible post-secondary education and training is essential to providing hope and opportunity to the disadvantaged (Lemann, 1995; Jansson, 1988).

In California, the community colleges are to serve a functional role in providing a vehicle for social and economic mobility. In reform legislation passed in the late 1980's the state legislature and various study groups gave particular attention to the role of the community colleges in providing a leveling influence on access to economic equality for individuals and groups . The reform legislation (Assembly Bill 1725), was recently amended to include economic development as part of the triad of the community college

mission that also included lower division preparation and occupational training. In recently adopted welfare reform legislation, the community colleges are to serve as the primary providers of training and education for welfare recipients. The community colleges are to serve everyone, from students intending to transfer to senior institutions, to those required to attend as part of a welfare or citizenship obligation.

The Social Role and the Underprepared Student: Pressures on the Colleges

The social role of the community college is an essential part of understanding the debates and arguments over the use of standardized tests as selection and classification instruments to place students. Fundamental to this debate is the problem of the underprepared student. According to Astin (1985), one of the most formidable tasks facing American higher education is the challenge of effectively serving and developing the talents of the underprepared student. He stresses that even the term, "underprepared" lacks precision or an operational definition. Astin points out that "Most academics seem to use the term for students who are not presently able to do 'college-level' work (another relative term) or who function at a level below that of most other students at the institution" (p. 105). Astin cites evidence of increasing numbers of students needing remediation, tutoring, and special educational support services.

As greater numbers of underprepared students present themselves for enrollment, community colleges will witness increased demand to offer courses and services to meet their needs for remediation, tutoring, and special assistance (Cohen and Brawer, 1996). However the definition of what an "underprepared" student is can vary as much as the institutions who deal with students of divergent abilities and preparation.

The community colleges are viewed by some policymakers and scholars as exerting a leveling influence on access to opportunity and wealth. Community colleges are

viewed as the primary system that offers a chance to economically disadvantaged persons in the vast system of American higher education and post-secondary training. The concept of open access represents more than low cost, physical proximity, and no requirements for admission to the colleges. Open access suggests that economic and social mobility is hindered by an attenuation or loss of access to the programs offered by the community colleges in any form. To critics of testing and tracking, mandatory testing and prerequisite skill levels for college entry represent a diminution of access for all to the community colleges.

Debates over Access versus Quality: The Antecedents of Reform

Unlimited enrollment growth in the California Community Colleges during the last two decades eventually collided with limited state funding. California voters grew weary of what was generally perceived as profligate spending by the state government and passed referendums that severely curtailed the growth in tax receipts that helped to pay for expanded community college classes, program, curriculum, and campuses. After witnessing the profound shift in the behavior of the voters, legislators themselves soon became champions of lower taxes and decreased government spending and passed additional legislation limiting the growth of government programs, particularly those in education. For example, enrollment limits or "caps" were soon imposed on all community colleges where each college would only be funded up to a predetermined number of students. This was intended to limit growth in college programs and enrollments. Following the passage of property tax reform in the late 1970's, local governing boards were virtually unable to raise taxes to pay for expanding programs and increasing costs. The burden for funding the community colleges shifted to the state. Loss of local control led to profound changes in funding formulas, acceptable courses or programs, and greater

accountability. Legislative and voter patience with unlimited open-access and community studies courses had worn thin.

The 1980's: A Decade of Community College Educational Reform

In the mid-1980's, property tax reform, a general economic downturn in the state, and the sudden shift of funding and local control to the state government, led to a re-examination of the policy of unrestricted open-access. Legislators and other elected officials grew increasingly restive with funding policies that encouraged enrollment growth in practically any program or course that the colleges wanted to offer. As funding declined and legislative oversight increased, open access policies were brought into question. Open access and program growth, once viewed as the open-door to opportunity, economic betterment, cultural renewal, and lifelong learning, was increasingly viewed as harming educational quality and detrimental to academic standards.

It was during the 1980's that the California legislature enacted sweeping reform of the community college mission, governance, and student enrollment and matriculation policies. State lawmakers and officials, and statewide bodies such as the community college State Academic Senate and others believed that reforms would result in a re-emphasis of the lower division function, greater professional autonomy of college faculty and staff, and improved student outcomes in transfer, graduation, and occupational placement.

Accountability and evaluation of college programs and services were also included in community college reforms. Through the collecting and reporting of accountability indicators such as student persistence, retention, graduation, transfer, and occupational placement, policymakers also hoped that the colleges would become more effective in monitoring and reporting on local efforts to reverse the perceived downward trend in

academic standards and student outcomes. Prescriptive reform, it was believed, would show the colleges how to make things better. Accountability, it was hoped, would provide an incentive to get them to do it.

Thus did the access versus excellence debate renew itself in the mid-1980's. To some observers, the successful accomplishment of the access mission helped to fuel current debates on quality (Deegan & Tillery, 1985). As open-access institutions the community colleges have witnessed tremendous growth over the past ten years in the demand for basic skills courses in English, mathematics and English language instruction. With the perception of declining academic skills, primarily in the areas of basic reading, writing, and mathematics, educators at the post-secondary level reported a sharp rise in the number of students requiring basic skills instruction and remediation (Cohen, 1987). These observations, combined with a decline in student transfer, graduation, and persistence rates, led to a thorough examination of the mission and purposes of the two-year colleges in California in the late 1980's. Reforms in governance, mission, and accountability soon followed, among these was the Seymour-Campbell Matriculation Act of 1986 (AB-3, 1986).

Background of the Matriculation Reform

Although several reforms were passed by the California state legislature during the 1980's, the Matriculation Act is of particular interest to this study. This Act mandated that California community colleges design and implement a Matriculation program to improve student outcomes. This legislation was intended to improve student outcomes in the two-year colleges in California by more clearly delineating the responsibilities of students as participants in the educational process and providing enhanced support services to help students meet those responsibilities.

The Matriculation Act was passed in 1986 after a flurry of national and state reports had been released describing a steady decline in academic standards and performance of the two-year colleges and education in general. Paralleling the tone of reports such as "A Nation at Risk" (National Commission on Excellence in Education, 1983), reports prepared by the Academic Senate for the California Community Colleges suggested that academic standards were declining. It is not clear from senate reports if the term "academic standards" referred to the entering abilities of students, or the performance demands of the colleges. However in the view of the Academic Senate, a perceived decline in academic standards was equated with a similar decay in the reputation and quality of community college educational programs.

The State Academic Senate decided to focus on the problem of declining student abilities as an indication of lowered academic standards. the perceived decline in the academic quality of the two year colleges was studied and debated at the state level in legislative hearings and through several taskforces. These studies and debates were eventually transformed into calls for accountability and reform in the governance, structure, mission, and matriculation practices of the community colleges. Two major intentions of these reforms were to delimit the community college mission and make the matriculation process a more engaging and intensive process for students.

The legislative version of Matriculation can be traced to the fall of 1982 when the State Academic Senate for the California Community Colleges adopted a resolution entitled, "The Matriculated Student" (Alkin and Freeman, 1992). This resolution stated that community college students seeking a degree or certificate should be designated as "matriculated," and be held to certain requirements in order to maintain that status. In June, 1983, the Board of Governors for the California Community College system approved a model for matriculation services presented to them by state officials. This model was then

field-tested at 16 California community colleges. Favorable reviews by state and local college officials resulted in the introduction of the Seymour-Campbell Matriculation Act in 1986.

Undergirding the design of the Matriculation reform was the belief that making the college matriculation process more structured and intensive would engage students in their own educational planning and help them to set realistic educational goals. Assessment and placement testing was intended to aid in this process by providing information to the student about their aptitude to read, write, and compute at the collegiate level. Testing and placement, in conjunction with enhanced student support services, was supposed to lead to improved teaching and learning in the community colleges. Improved teaching and learning would lead to improvements in college performance indicators. Thus community college reforms were intended to lead to improvements in the transfer, graduation, job placement, and retention rates of students.

Assumptions Underlying Mandatory Placement Testing:

Through the increased use of testing and mandatory placement and prerequisite courses, advocates for the Matriculation reform believed that the colleges would improve their effectiveness and hence student outcomes. Mandatory placement through test scores and prerequisites was to yield a direct benefit to the student, faculty, and college as a whole. There was a belief that creating homogeneous student groupings using test score ranges and cutting scores would improve student outcomes and success at various levels in the curriculum (Kingan & Alfred, 1993; Boggs, 1984).

Another argument advanced in favor of expanded use of testing was that students are less likely to drop out if they were placed in courses that were not too challenging for them. The retention of students who had the required score or prerequisite would improve (Hughes and Nelson, 1991). Students would be less likely to drop out due to frustration

with a class that moved too slowly because of the presence of underprepared students. There would be less disruption during the enrollment period as students without the required ability level would not be allowed to fill seats in crowded classrooms needed by students who had the required test score or course prerequisite.

Matriculation reform was also intended to reduce college costs. Grouping individuals by ability level was supposed to more efficiently focus instruction in the college level courses by alleviating the need for faculty to remediate poorly prepared students.

To some critics of standardized placement testing there were drawbacks to this approach. They viewed college practices that sorted and placed students into pre-collegiate or collegiate level courses with suspicion. Critics believed that the use of test scores, mandatory prerequisites, and counseling to place students resembled student tracking, ability grouping, and a restriction of access. From the perspective of the testing critics, using test scores to direct students to levels in the curriculum or for program eligibility represented a type of test tyranny that justified exclusion, and systematically and unfairly throttled student aspiration. (Kingan & Alfred, 1993).

Critics of the use of testing were also joined by some academics and policymakers who view the community colleges as an institution that maintained social stratification and solidified class differences (Birenbaum, 1986). Critics of testing and placement held compatible views with those who believed that the community colleges had not fulfilled their economic and social obligation of uplifting entire groups of people. According to testing critics, placement testing, sorting, ability grouping, and realistic goal setting for students were the mechanisms used to throttle student goals for transfer and further higher education opportunities (Brint & Karabel, 1985; Zwerling, 1986).

Demands for Test Accountability Evidence: The Institutional Response

As a growing number of state legislatures require pre-admissions testing, there is a growing interest in examining the outcomes of such policies.(Wattenbarger and McLeod, 1980). In a climate that increasingly emphasizes accountability and assessment, describing and monitoring the effectiveness of placement practices assumes greater importance (Bers, 1989). The credibility and fairness of an assessment system relies in large measure on evidence that placement testing produces better and more equitable outcomes for students (Bers, 1989).

Community demands for evidence of test validity, multiple measures, and fairness placed new burdens on the research capabilities of state agencies and community colleges. Institutional research in the community colleges had traditionally not been well-funded nor staffed (Cohen, 1996). Requirements for test validation studies, disparate impact studies (the effects of testing on particular groups, such as ethnic and linguistic minorities), and the effects of ability grouping on certain student populations placed new demands on a poorly developed or fledgling research function in the community colleges. However, since the passage of the Matriculation Act, many community colleges have established a research and evaluation function (Alkin and Freeman, 1992).

Of those colleges that had established institutional research offices, many of the institutional research staff did not have training in testing and psychometrics. Sheldon (1970) observed several methodological and procedural problems associated with attempts by community college staff to predict academic success from standardized tests. Among the problems identified by Sheldon were a misunderstanding of the concept of validity, misinterpretation of correlation coefficients, misuse and misunderstanding of statistical significance, and a general failure to recognize that test items are nothing more or less than a sample of behavior. Sheldon also pointed out that standardized tests add little to the

prediction of academic success beyond that contributed by high school grade point average, thus he questioned their presence on the community college campus. With the advent of new and stringent testing regulations, there was much confusion in the community colleges about how to proceed to gather and report the required evidence. Colleges embarking on testing programs appeared poorly prepared to conduct predictive validity and other studies to satisfy the demands of state agencies and court decisions.

In California, there appears to be a legislative ambivalence toward testing in the open-access college. On the one hand, pre-admissions testing is mandated, on the other hand there are a number of state regulations designed to curb what is perceived as the excesses of testing. These regulations regarding the use of testing did not come about solely in response to lawsuits brought by advocacy groups and testing critics. Distrust of testing and the motives and competence of test users has a long history.

The Testing Debate: A Continuing Saga

The recent and continuing debate over the value and use of skills testing is not new. Disagreements over the value of the inferences made by test users regarding the abilities of test takers can be found in historical writings from ancient China (Miyazaki, 1981), to more contemporary debates over the conclusions by the authors of The Bell Curve (Herrnstein and Murray , 1992). The higher the stakes involved in a score-derived (e.g., test score) inference about an individual, the more vociferous the debate. If tests are used to allocate opportunity to prestigious universities, elite training programs, military services and attendant benefits, well-paying jobs, or to college preparatory programs in high school, the views on testing become even more polar.

Increasingly, members of groups traditionally under-represented in higher education and in management in the private sector are attempting to enter higher education

institutions and private companies. If testing is seen as reproducing in the elite universities and large corporations the class differences and stratification in society, then the debate over the use of standardized becomes more heated. The debate over the fairness of allocating opportunity using tests often includes overtones of race and class distinctions (Collins, 1979).

Testing as an indicator of aptitude and achievement has the ability to provide fodder and sustenance both to those who believe in the need for testing, and those who decry the use of testing. In the middle are those who think testing while not perfect, is the best alternative available (Jencks, 1989).

The testing debate, although particularly acute today, has been steadily growing over the last 100 years. Throughout this century there has been a growing interest and use of score-based inferences to select, sift, sort, screen, classify, and group individuals for education, training, the military, or the workplace.

Ability Grouping

Many high-stakes decisions affecting individuals are made on the basis of test scores. For example, test score ranges are often used in educational settings to track and place students in "ability-groups." (Tyler, 1974; Oakes, 1985). As noted by Oakes, ability grouping in education has a history several decades long. Not long after their use by the military services in 1917, mental tests were adopted by school officials and teachers in the 1920's as a method to group individuals for instruction. Schools began to adopt a three-tiered grouping plan with the those of highest mental ability in the top tier, the lowest in the bottom tier, with the mid-range students in the middle tier. Although there has been tremendous ebb and flow in the use and popularity of ability grouping in education over the

last 80 years, beliefs in the efficacy and need for ability grouping persist. This despite questions of its utility and fairness (Oakes, 1985, Tyler, 1974, Goodlad, 1969).

According to Oakes, there are "mountains of evidence indicating that homogenous grouping doesn't help anyone learn better." Oakes cites several studies conducted over the last half-century that suggest that ..."no one group of students has been found to benefit consistently from being in a homogenous group." (1985. p. 7). In higher education, a review by Willingham (1974) also suggested that the evidence for the usefulness of grouping strategies was inconclusive. Although there are many who support the idea of creating homogenous student groupings using placement testing for instructional purposes in the community colleges (Bers, 1997; Bray, 1987; Boggs, 1984).

The logic and approaches of ability grouping found in the primary and secondary schools may also be found in the community colleges that use placement tests to determine achievement and group individuals according to test score ranges. According to Oakes (1985), ability grouping and tracking are based on beliefs and assumptions about how such practices improve instruction. For example, it is believed that students learn better when they are grouped with other students who are like them academically. Brighter students will not be held back by duller students. Another assumption is that slower students develop a more positive outlook by being grouped with others. A third assumption is that the placement process used to separate students into groups both accurately and fairly reflects past achievements and native abilities. Similar arguments and rationale regarding the logic and rationale of ability grouping can be found in legislative testimony, State Community College Academic Senate reports, and in conversations with many faculty leaders in community colleges. With the growth of these practices in higher education, it appears that ability grouping is no longer just for kids

The use of tracking as a method to accommodate individual differences was not limited to the elementary and secondary schools. Higher education has also relied on similar approaches to accommodate and group individuals of varying ability (Browne-Miller, 1995). Kingan & Alfred (1993) identify two distinct methods of student tracking in their research. In the first method, all students follow a similar curricular program, but each student is grouped within classes at a particular ability or aptitude level. The second approach tracks students into different curricular programs according to ability. They contend that most community colleges use the second method to sort and track students. Here, the community college separates students into supposedly homogenous instructional groups based on inferences about individual ability or aptitude. This observation on the predominant tracking approach in the two year college is similar to what Clark (1960), found in his seminal case study of a junior college in California. Although Clark noted the tendency of community colleges to redirect student aspiration, he found it to be more of a tacit process. Advocates for increased use of testing and placement would like this function of the community colleges made more explicit.

Mandatory testing and placement was also supposed to benefit the well-prepared and high achieving students. For example, in conversations and meetings with this investigator, advocates for testing and mandatory placement suggested that grouping the highly motivated and prepared students would improve the retention of this group (Armstrong, 1989). This view is echoed by testing advocates who also suggest that the high ability students will be less likely to drop out due to frustration with a class that moves too slowly because of the presence of low achieving or underprepared students in the classroom (Beatty-Guenter, 1992).

Creating homogeneous student groupings using test score ranges is problematic. The problem is compounded by the use of measures and predictors that are only analogous

to the desired outcome to classify people. For example, using a multiple choice test to determine eligibility for a writing course may not predict well because of the lack of correspondence between a course objective such as essay writing, and choosing the one best answer on a test form. Some scholars assert that ability grouping often harms disadvantaged students by inaccurately classifying them as highly unlikely to succeed in a particular course or program of study (Henriksen, 1995; Sticht, 1987; Oakes, 1985). This may be due to their lack of familiarity with standardized assessment tests.

From a broader perspective, Browne-Miller (1995) notes that the three-tiered system of higher education in California might be viewed as a three-level ability group, similar to the ability grouping strategies identified by Tyler (1974) in the elementary schools. On the basis of test scores, high school grade point average and activities, and some additional criteria, students are classified and selected for the public higher education system in California. The highest in ability (the top 12.5% of high school graduates) are eligible for the UC, those in the top one-third are eligible for the California State University, while the community colleges are open to any adult in the state regardless of academic background, education, or goal. These divisions might be viewed as reflecting beliefs about the distribution of general mental ability in the high school graduate population. Within the institutions, tests are used to determine eligibility for certain courses, to measure progress, and to certify mastery of course content. If a student goes on to graduate school, then much of this same information is used to determine eligibility and selection into various programs.

Testing and Prerequisite Skill Levels: The Potential for Misclassification

The growing reliance on standardized tests to infer ability and sort students in the open access college is troubling to those who view testing with suspicion. To critics of

testing in the community college, as in other educational or training settings, perhaps the most egregious aspect of standardized testing is the potential for misclassification. With the understanding that tests are not perfect indicators of actual performance in school, training, or on the job, it is expected that many people who could perform well will fail or score below the score set as a minimum for entrance. Thus using a test score-based inference may reject many who are capable of succeeding.

Interpretation of test scores is also a potential problem. Sheldon (1970) observed that community college staff did not seem well prepared to understand the limits and appropriate use of test scores. Although many California community colleges now have a research function, test validation is only one of several activities they are asked to conduct (Cohen and Brawer, 1996). This problem of interpreting test scores or inferred prerequisite knowledge is essential to understanding the role of testing and its importance in acting either as a barrier to opportunity, or as a gateway to success in the community college.

Many of the same complaints about the poor preparation levels of young people heard today were voiced many decades ago in the armed services. The use of group testing enabled entire generations of young men (at that time) to be described and ranked along a scale. Those ranked highly were presumed more able, hence more desirable for training. Those ranking lower were deemed suitable for lower level positions, but were seen as difficult to teach or train. A review of a pertinent example of classification will help to place the current debate on testing in the community colleges in an historical context. The following discussion suggests that while the underlying beliefs about mental ability have changed dramatically since 1917, the strategies for accommodating divergent mental abilities have remained remarkably constant. The armed services' experience helps to provide a historical example.

Mental Classification in the Armed Services: A Case in Point

The first mental tests designed to be used for mass, group testing were developed by psychologists for the U.S. Army in 1917. The group tests were modeled after intelligence tests designed for individual use in one-on-one assessment. As noted by Sticht (1992), both tests were based on the theoretical position that intelligence was an inherited trait. It was further assumed that native intelligence was being assessed. Two tests were developed, the Army Alpha and Beta tests. The Army Alpha test was intended for literate recruits. It was a paper and pencil test battery that included tests ranging from simple or complex oral directions to the knowledge of arithmetic and a wide variety of vocabulary and general information items. It also called for common sense and practical judgment, and analogical reasoning. Scores for the eight subtests were combined into a composite score. This score was assigned ranges, and these ranges were assigned letter grades. A letter grade of A signified superior intelligence, B, C+, and C signified average intelligence, while C-, D, or D- indicated inferior intelligence. The letter grade became the person's mental category and was viewed as a general indicator of native intelligence.

The Beta test battery intended for illiterates and non-English speakers, also used a number of subtests. Each of the subtests was administered in pantomime by the test examiner and aide (Yoakum and Yerkes, 1921). The examinees marked their responses on paper using pencils, but they were not required to use written language. As with Army Alpha, the Beta subtest scores were combined into a composite score. The composite score was viewed as a measure of native intelligence.

Scores from the Army Alpha and Beta tests had many intended uses. One use was to classify recruits according to their mental ability and assisting with assignment to jobs that suited their demonstrated mental capacity. Another intended use was to supply a mental rating or common metric for each recruit to assist personnel officers in building

organizations of equal mental strength. This was to lessen the likelihood that any particular section of the Army would receive a disproportionate share of mentally inferior soldiers. It was intended that the tests would assist in discovering recruits who were mentally inferior and thus should be recommended for discharge, "development battalions," labor organizations, or regular military training.

Although the Army Alpha and Beta tests were not widely used after the war, their use in the armed services guided thinking about intelligence and aptitude that continues today. For example, the tests introduced the concept of mental categories and the notion that personnel could be assessed with regard to their mental abilities using paper - and - pencil tests. Personnel thus assessed could then be reliably and accurately assigned to a mental category. This notion in large part still exists today (Sticht, 1992).

As in other sectors of society, the use of tests in the armed services as screening or sorting devices often fluctuates in response to the available supply and demand for personnel or students. For example, the Armed Forces General Classification Test (AGCT) was not used as a screening device until after the Second World War when manpower needs had declined. In 1948 the U.S. Congress passed the Selective Service Act which designated an entire mental category, (Category V) as unfit for military service and they were thereafter excluded from enlistment by statute. According to Sticht (1992), this excluded some eight percent of the young, white, male population from eligibility for service. This was the first time that a statutory mental standard was set for enlistment in the armed services (Sticht, 1992).

According to Sticht, (1987), many of the beliefs about the mental quality of lower scoring recruits seem premised on perception and bias. For example, during the war, increasing manpower needs compelled the military to progressively lower the required mental standards for service. It was during this time that many commanders began to

complain about the low levels of aptitude and mental ability among lower scorers on the Armed Forces Qualification Test (AFQT). They complained that training time was prolonged, and that these lower ability recruits were slowing the progress of more able recruits. In addition the combat readiness of various units was being hampered by the difficulty in training these young men. Faced with these complaints by commanding officers on the one hand, and the press of manpower needs on the other, the Department of the Army came up with a solution:

> The Department of the Army then arbitrarily decreed that the top half of that category (a low mental category) would henceforth be classified in the next higher category. Commanders practically ceased their complaints, although they were getting the same number of low quality of men as before--but now only half as many were designated as being in the lowest mental category. (U.S. Department of the Army, 1965, p. 4).

Perhaps the stigma attached to the various categories led many to complain about the low mental quality of new recruits. The anecdotal evidence cited above highlights the importance of labels in determining the perception of one's fitness for a course, program, or job. In this case, top military officials were aware of the redesignation of the lowest ability men into the next higher category. There have been other times when lower ability personnel were admitted into the nation's armed forces when officials were unaware that this was occurring. A review of this experience was instructive in the formulation, rationale, and design for this investigation.

In 1976, the Armed Services Vocational Aptitude Battery (ASVAB) was introduced as the official test used by all services. The ASVAB combined the Armed Forces Qualification Test (AFQT) and special aptitude tests into one battery that included ten subtests. These tests included word knowledge, paragraph comprehension, numerical operations, arithmetic reasoning, mathematical knowledge, coding speed, general science, mechanical comprehension, electronics information, and automotive shop information

(Sticht, et. al, 1987). However, the introduction of the ASVAB did not proceed flawlessly from the military's point of view. Its introduction produced an unintended opportunity to examine the problem of misclassification under almost experimental conditions.

Between 1976 and 1980, tables used to calibrate the ASVAB were accidentally misnormed. As a result the armed services unknowingly enlisted hundreds of thousands of category IV recruits far beyond the established quotas. According to Sticht et. al. (1987), this provided a "natural experiment" to discover how well low-aptitude individuals perform when their measured aptitude levels are unknown and they are thought to be of higher aptitude. Regarding the misnorming, Sinaiko (1988) (a noted military human resource specialist) observed that: "The misnorming was the largest and perhaps, only triple blind experiment ever done: no one--soldiers, officers, or military manpower executives--knew that it had occurred for over four years. Amazing!"

The effect of the miscalibration of the ASVAB was to assign higher percentile aptitude scores to examinees than they would have received had the norming tables been correct. For example, it was believed that only 9 percent of recruits were in mental category IV (10-30 percentile) of the distribution. However, recalibration of the corrected ASVAB tables showed that 46 percent of inductees were in this lower ability group. Only the lowest 9 percent of the total number of applicants to the armed services had been denied enlistment during those years based on their ASVAB scores (Shields and Grafton, 1983). As of October 1, 1980, the Department of Defense had inadvertently accepted over 300,000 "miscalibrated" recruits (Sticht, 1992).

Performance of the Miscalibrated

Discovery of the test calibration error and the enlistment of hundreds of thousands potentially non-eligible personnel did result in several retrospective studies. These were

again analyzed and reported in summary form by Sticht, et. al. (1987) and by Sticht, (1992). Sticht and his associates found that, overall, the miscalibrated recruits performed slightly worse than the average-aptitude control group in such areas as promotion, eligibility for reenlistment, and attrition. However the differences he noted were not large and in several cases the miscalibrated soldiers performed as well as or even better than the average-aptitude control group. Where there were larger differences these were often due to the particular personnel policies of the varying branches of the services. For example, promotion in the navy or air force relies more on paper and pencil tests for advancement. In the army however, supervisor ratings of performance are used primarily for promotion and to determine eligibility for reenlistment in the all volunteer force. As Sticht (1992) observed, "In general, whenever lower aptitude personnel are placed in a test-taking situation, they do more poorly than when their work is evaluated by supervisor ratings." (p. 63).

The performance of these "stealth fighters" as the miscalibrated have been called (Sticht, 1997), calls forward the question of the criterion used to judge performance and against which we validate instruments, skill levels, and placement practices. It appears that a valuable lesson from the ASVAB miscalibration is that ratings of performance by a human judge can vary tremendously and may reflect other factors pertinent to success. Selecting a sound and reliable criterion becomes of paramount importance in judging the value of any instrument or practice used to allocate opportunity. A focus on the criterion used to validate placement tests is a line of research in the present investigation.

An important opportunity was gained in reviewing the experience of the armed forces when persons thought ineligible or unable to succeed were given the opportunity to try. Similarly, this investigation will take advantage of a period when college placement rules were temporarily relaxed. This allowed many students to ignore placement advice

and attempt courses or programs which they otherwise would have been prevented from trying. This is a unique opportunity to analyze and potentially improve placement practices in the two year college setting.

Testing and Classification in Education: A Historical Overview

Making inferences from a test about a person's ability to succeed in some task at some time in the future has a long history in education. One of the first applications of tests for this purpose was conducted by Alfred Binet in 1904. Binet, an early developer of intelligence tests, was asked by authorities in Paris to help them devise a method to identify bright children and to remove non-learning children. Once identified, these non-learning children could be sent to another school, program, or class where they could be taught from a simplified curriculum. It was believed by teachers and school officials that the slower learning children were impeding the learning of the bright children by requiring excessive amounts of attention and teacher time. They also believed that the slower children could better profit from slower paced, remedial instruction in another school. School officials agreed to assist the teachers in this but wanted an objective method to sort the students rather than using teacher judgment. School officials worried that if teachers made the decision then they would also attempt to rid themselves of disruptive, but erstwhile bright students, rather than the slow learners. Teacher judgment was also not used because some officials feared that the slower children with pleasant personalities favored by the teacher might not be rejected.

Binet's solution was a composite scale comprising several different types of items demanding a variety of tasks. After examining the test items he developed and the correlations of the scores on these items with other indices of general ability, Binet arrived at his definition of intelligence. He defined intelligence as "...the tendency to take and

maintain a definite direction; the capacity to make adaptations for the purpose of attaining a desired end; and the power of auto criticism" (translation by Terman, 1916, p. 45).

Binet arrived at his definition of intelligence or general mental ability through trial and error. He noted that those who were best in judgment were also better in vocabulary, attention to detail, and logic for example. These separate indicators of ability seemed to be correlated. This helped him to formulate his description of a general mental ability. His original scale formed the basis of several mental tests that were developed soon thereafter. Many of the items developed for these earliest mental tests exist in some form today.

Binet's discovery seemed to come at an opportune time. There was a growing and generally accepted belief in the almost omnipotent powers of science to address the most vexing problems of society. According to Cronbach (1990), at the turn of the century there was a great demand for objective methods of sorting and classifying individuals. Binet's scale appeared impartial and objective and was hailed as another triumph of science in the identification of the gifted from the dull. An objective and scientific method of identifying intelligence, aptitude, and achievement had been found. The possibilities were only beginning to be realized.

Mental Testing as a Reflection of Societal Values

The development and subsequent popularity of mental testing and its application in identifying the most talented also mirrored societal beliefs about ability. Darwinian concepts on evolution and distribution of talents and traits enabling survival were also reflected in beliefs about how intelligence was distributed. These beliefs also coincided with meritocratic ideals of American democracy. The notion that the "cream would rise" from among the masses reflected Darwinian concepts of evolution and survival of the fittest

and that American capitalism would reward the high achiever, regardless of previous social rank.

There was a belief that intelligence was largely fixed and immutable. Increasingly, tests came to be used to identify those with the greatest capacity. Thus if some individuals were more fit from an evolutionary perspective, then this group would be one most likely to benefit from education. Although many of these beliefs are far less prevalent today, some argue that these early beliefs about the nature and distribution of intelligence, mental fitness, and fixed mental capacity are still reflected in actions and decisions made by schools, colleges, and employers. This is reflected in current views about ability grouping, tracking, and ability to benefit from instruction. In most cases decisions about grouping, tracking, or ability to benefit is based on an inference derived from a test score.

The Use of Testing as a Sorting and Classification Device

Almost from their inception, tests were used as an administratively convenient and inexpensive tool to help solve an array of difficult allocative problems in education and the workplace (Lemann, 1995). The mass immigration of hundreds of thousands of persons from countries in eastern, central, and southern Europe in the early 1900's resulted in a tremendous diversity of background and educational level that was largely unfamiliar to American officials. With a shortage rather than a surfeit of well educated, well trained recruits, the military services were compelled to develop a means to identify, sort, screen, and classify recruits according to ability and aptitude. In contrast with earlier periods in which the military services could screen and deny entry to applicants, in this instance the military was faced with the need to rapidly expand from its small peacetime size, quickly deploy personnel, identify those with leadership ability, accommodate tremendous diversity in preparation and cultural backgrounds, and adjust to the needs of an increasingly

technological and mechanized work setting (Yoakum & Yerkes, 1920). In comparison with the rhetoric heard today regarding the increasing technological demands of work and diversity in background and preparation of the workforce, the rationale and reasons for testing in 1917 sound remarkably contemporary.

Following World War I, standardized tests moved into the classroom and workplace. Schools, colleges, and large industrial organizations were seen as the proper civilian settings for the initiation and development of testing. Educational institutions and the personnel departments of industrial organizations were asked to help select and sort applicants. The success of mental testing in identifying the ability of thousands of recruits to absorb training in the military suggested to civilian leaders that similar tests could be developed for the civilian sector.

As in the armed services, tests were used originally in the schools primarily as selecting and sorting tools rather than as educational or diagnostic tools. The use of testing in educational settings was based upon the psychology of individual differences rather than upon the psychology of learning (Tyler, 1974).

Testing in Higher Education: A Response to Demographic and Policy Shifts

Within higher education institutions, there have been many approaches and shifts in the methods used to accommodate individual differences among students. What influences the adoption of various approaches can be politically or policy related, based on demographics, or some combination of the two. Views of how to allocate opportunity in higher education may have their basis in politics or interpretations of events. For example, during the Cold War, there was a wave of interest in promoting science and mathematics education following the launching of Sputnik in the late 1950's. Public policy became increasingly focused on the relatively small proportion of the age group who demonstrated

academic progress in the sciences. This was seen both as a prudent use of resources, and as an investment in national security. The space race between the U.S. and Soviet Union launched by Sputnik also launched a knowledge and technology race between the two nations. The nation's most selective colleges and universities were chosen to help relieve the national anxiety over a perceived technological knowledge "gap," that paralleled the much ballyhooed missile gap at issue in the 1960 presidential race.

Demographic pressures have also affected attitudes and approaches within higher education toward individual differences. The large number of high school graduates in the mid-1960's focused attention on increasingly selective admissions. Willingham (1974) suggests that one result of the surfeit of high school graduates was the stratification of higher education institutions with respect to academic ability. Educational stratification within higher education was also noted by Browne-Miller (1992).

Following the "baby-boom" surge of high school enrollments, higher education began to face declining enrollment in the late 1960's and early 1970's. Higher education institutions responded by placing greater emphasis on access to higher education for all types of students. As noted earlier in this section, the armed services have also periodically redefined the notion of quality personnel by adjusting minimum requirements for entry in response to manpower supply and demand. Within higher education entry standards also appear to also adjust depending on the perceived supply of available students.

Changing social values about fairness and equity and a desire to address social inequality also affected approaches to accommodating students of varying academic abilities. The decline in high school graduates observed in the late 1960's (Willingham, 1974) was also coupled with increasing priority to enlist the nation's colleges and universities in the War on Poverty. New social priorities coincided with the access emphasis in the late 1960's and focused on increasing opportunities to attend college for

students historically under-represented in higher education. The shift in emphasis to access and recruitment of disadvantaged students into higher education placed testing and the uses of testing in higher education in a different light.

As noted earlier in this chapter, unrestricted open-access in the community colleges gradually fell into disfavor in the 1980's. As a result, there were renewed calls for testing as a device to classify learners according to aptitude or ability. For example, passage of the federal Ability to Benefit legislation mandated that students seeking federal financial aid and not possessing a high school diploma must have achieved a minimum score on a standardized test approved by the US Department of Education. This was intended to identify those who could profit from instruction. Presumably, persons scoring below the designated cutting score are not able to benefit from college level instruction. In 1996, the Florida legislature passed a law that mandated higher cutting scores for students attempting to enter college level courses. Analyses of these new cutting scores suggest that 72% of new students would place into pre-collegiate remedial courses. This is almost a 25% increase above the proportion referred to remediation under the previous cutting scores (Ignash, 1997). This policy is intended to limit the costs of remedial education by limiting space to those thought to be higher aptitude.

Testing and Mandatory Prerequisite Skill Levels in the Community College: Two Sociological Perspectives

Why are the debates over the use of placement testing in the community college so heated and prolonged in contemporary society? Part of the answer to this question lies in conflicting views over the role played by the open-access college in American society. Debates over the role of testing are part of a larger debate over the role of education in social stratification that has been going on among sociologists for several years now.

Particular attention in this debate has been paid to the role of the community colleges in perpetuating or mitigating inequality.

Functional Theory

On one side of this debate are the functional theorists. Functional theorists argue that social stratification serves a useful purpose. It assures that society will have sufficient human resources at all levels. Division of labor is essential to a well functioning and efficient society and social stratification helps to ensure that there will be individuals distributed among the different positions needed by an increasingly technological society. Some positions require more skills, training, and knowledge than others. Educational stratification assists with the task of filling the most difficult positions with individuals of the best talent. Talented individuals must be motivated to fill important positions needed by the larger society. During periods of national mobilization and crisis, it is vital to the national security and economic well-being of the nation that a society uses its mental resources in the most efficient way possible. A system of unequal rewards helps to provide that motivation. Structured inequality with a system of rewards provides the motivation necessary for society to fill its most vital positions. Under the functionalist point of view, the community colleges are an important institution in the identification, selection, training, and placement of individuals into positions commensurate with their abilities. Talent is not equally distributed among all individuals. Rather, there is a distribution of talent resembling the bell curve. To the functional theorist, the community college has performed an important role in furthering American ideals and approaches to social and economic mobility. Functionalist beliefs about the value of the open access college can be found in mission statements, public pronouncements by education leaders, and legislative discussions about the role of the community colleges.

Conflict Theory

In contrast to the functional theorists, the conflict theorists maintain that social stratification is inherently unequal. They have an inherent distrust of the bell curve. Conflict theory holds that social stratification does not provide the motivation necessary nor does it ensure that the most critical positions in society will be filled by the most able. Instead they maintain the class structure ensures that the powerful and wealthy will maintain their relative positions through the imposition of a system that imposes and recognizes a largely artificial privilege (Collins, 1971). Community colleges perpetuate this social inequality based upon class privilege (Brint & Karabel, 1989; Karabel, 1986). They suggest that student's receive education roughly corresponding to their parent's position in society. The community college with its open-door admissions policy are seen as merely symbolic gestures to ameliorate the hidden injuries of class (Zwerling, 1986). To the conflict theorist, the toll of the Bell Curve does not resonate well. Rather view the bell curve as exacting a heavy toll on aspiration.

Conflict theory comports well with the beliefs of testing critics. They suggest that growth in the use of standardized testing to sort persons according to ability to absorb learning at various levels in the curriculum is another example of how the community colleges act to throttle the aspirations of the poor and disenfranchised.

Functional theorists on the other hand maintain that there is a relationship between academic ability and program eligibility that is independent of socioeconomic status. Clark (1960) described this phenomena and introduced the "cooling out" function of the open-access college. The colleges do not actively thwart aspiration said Clark, rather through a complex and subtle process they act to help the low achieving student rethink and redefine their educational goal. This is done in a variety of ways; through skills testing, counseling,

and special remedial programs. He suggests that democratic societies need to motivate individuals to strive to higher ends, but they also need methods and institutions to placate the low achieving or marginal student to sustain motivation in the face of disappointment and to redirect anger or frustration. "Democracy encourages aspiration, and generous admission allows the student to carry his hopes into the school or principally the college. But there his desires run into the standards necessary for the integrity of programs and the training of competent workers. The college offers the opportunity to try, but the student's own ability and his accumulative record of performance finally insist that he be sorted out" (Clark, 1962, p. 80).

Thus functional theorists have much in common with the proponents of testing in education and in open-access institutions like the community colleges. They suggest that testing and ability grouping students into remedial and non-remedial tracks improves the learning outcomes for students by placing them at levels in curriculum that will enhance their learning and success (Bray, 1987; Boggs, 1984; Bers, 1997). It also serves to prevent the marginal student from aspiring to unrealistic expectations.

This ongoing testing debate creates an institutional ambivalence in educational and training institutions, particularly in those that are founded upon principles of open-access. In the last ten years the testing debate has become sharper and more pronounced. No longer just the stuff of educational and measurement journals, or a reference in an admissions policy of a university or college, discussions about testing and its effects have moved to a broader and increasingly public arena (Gifford, 1989).

Proponents of expanded use of skills testing ask that we consider a viable and fair alternative to standardized testing for screening and classification. In his conference remarks to the National Commission on Testing and Public Policy, Christopher Jencks (1989) suggested that although standardized placement tests are imperfect as indicators of

current achievement levels or predictors of future success, they are in many instances superior to available alternatives.

This chapter has described various aspects of the historical, social, and cultural debate surrounding testing in contemporary society. The relevance of this broad national ambivalence toward testing as it applies to the two year college was given particular attention. Testing is increasingly used to make decisions about access, ability and aptitude. This is particularly true as education budgets remain static or decline. With increasing pressure to serve greater numbers of underprepared students, policymakers and educators are looking to the metric provided by a test score to determine who can and should be allowed to enter college. This debate, fueled by historical and social beliefs about testing and stratification will probably continue in the years ahead.

CHAPTER THREE

REVIEW OF LITERATURE

A review of theories related to understanding predictive validity and the use of score-based inferences and student biographical variables in predicting outcomes are used to construct a framework for this investigation. Theory serves as a useful framework to arrange disparate observations about students into coherent relationships. The present investigation draws upon a multidisciplinary base of theories that focus on measurement, student biographical characteristics and predictive validity. Theory was used to derive variables for analysis, and to posit relations between these variables and student course performance outcomes.

Of particular interest and relevance for this study were measurement theories that focused on the prediction of performance and other inferences about individual aptitude derived from scores. This chapter describes the use of measurement theory to inform the analysis and interpretation of predictive validity for the research hypotheses presented in the first chapter.

Also useful to this investigation was the literature on student retention theory. A review of retention theory provided a basis for organizing student biographical characteristics into coherent groupings or categories. After a review of several student retention and college performance theories, most notably those of Tinto (1993 and 1982); Pascarella and Terenzini (1991); Bean, (1982); and Astin (1975); the scheme or rationale identified by Cross (1981) was chosen as the most suitable for modeling the characteristics of the study participants used in the present investigation. Using Cross' approach, participant biographical variables were grouped into situational and dispositional variables. The Cross scheme for grouping student variables was used to model the effect of

situational and dispositional characteristics on the course performance outcomes of final course grade and retention. This chapter describes the conceptual framework used to develop the course outcome performance model that included the use of standardized test scores and student biographical characteristics.

Conceptual Approach to Predicting Course Performance

Of the several measurement theories reviewed, two theories emerged as most useful to different aspects of the present investigation. Aptitude Treatment Interaction or Trait-Treatment Interaction (ATI or TTI) (Cronbach and Gleser,1965; Willingham, 1975), and Point-to-Point theory (Asher, 1974) emerged as the most useful for providing structure and an interpretive framework for this investigation. Point-to-Point theory emphasizes that the more features commonly found between a predictor and criterion space, the higher the observed validity. ATI emphasizes the interaction of student aptitude (test scores) and traits (biographical and dispositional characteristics) with alternate instructional treatments (instructor differences). Taken together, Point-to-Point and ATI theory assisted in the identification of variables and postulated hypotheses for empirical analysis.

In arriving at Point-to-Point and ATI theories to guide the present investigation, several theories were reviewed to assess the potential contribution of each theory to the questions posed in this study. The following section provides a summary of review of theory and identifies the contribution of each theory to the final model proposed for this study.

Classical Test Theory

Classical test score theory assumes that each person has a true score that can be distinguished from errors in measurement. However no measurement instruments is

perfect, such that the calibration of each individual's true score has some degree of error. The difference between the true score and the observed score results from measurement error (Kaplan and Sacuzzo, 1982). Another assumption of classical test theory is that errors of measurement are random and are therefore likely to cancel out for large groups of persons, repeated measures of the same person, or composites developed from multiple indicators for the same person at one time. It is generally assumed that the distribution of random errors will be the same for all persons tested. Classical test theory uses the standard deviation of errors, called the standard error of measurement, as the basic measure of error. The standard error of measurement reveals on average how much an observed score varies from the true score. As part of mandated validation studies, community colleges are required to report standard errors of measurement for any tests used (State Chancellor's Office, 1992).

An assumption of classical test theory is that the criterion for a test as well as the test is reliable. For the purposes of this investigation, Classical test theory provided little guidance in evaluating the predictive validity of score-based inferences. Unlike methods that focus on the internal properties of tests such as internal reliability, validity requires the gathering of data from external sources. Thus while the test may demonstrate internal consistency among items, it may tell us little about how an individual will perform in particular class or training program. Classical test theory also assumes that the criterion used to measure validity is reliable. However this assumption is not always met in the community colleges or other educational or training settings. Lack of reliability in the criterion, as well as in the test, may severely confound conclusions concerning the predictive validity of tests. Thus classical test theory, while valuable, may not be suitable for validation studies in educational settings where there are reasons to suspect that the criterion variable is unstable or unreliable (Willingham, 1974).

Common Method Variance

The measurement error postulated by classical test theory can act as a confound to predictive validity. Common Method Variance (Jensen, 1993) suggests that identifying a person's true score on a test can be threatened by maturation or different life experiences of test takers. This theory is instructive in interpreting criterion-related validity when individuals who are similar in ability achieve vastly disparate scores on a test. Common Method Variance posits that individuals who are equal in achievement may differ in their familiarity with a particular type of test format and that this affects the validity of the test regardless of the abilities that a person brings to the testing situation. Some of the confounds might include, test anxiety, test sophistication, test-taking habits, attitudes toward tests, test directions, noise level of the testing environment, language of the test, native language and cultural background of the examinee (Gifford, 1990). This can also contribute significant error to the accurate measurement of ability or aptitude.

The Application of Decision Theory to Test Validation

Decision Theory focuses on estimating the utility expected from alternate courses of action. As such, it is a fundamental principle in the logic of a testing and placement program. Cronbach and Gleser's (1965) adaptation and application of Decision Theory to educational processes and procedures such as placement, grouping, special instruction, tutoring, remediation, course exemptions, and honors programs recognized that these procedures represent alternate treatments. Selecting and grouping students using cutting scores on a placement test assumes that some students will profit from certain interactions at a higher level than will other students. Course placement in the community college appears to be based on such beliefs, although it is seldom stated in these terms.

Proponents of mandatory testing and placement seem to implicitly base much of their arguments on notions that arise from Decision Theory. Decision Theory suggests that students will demonstrate higher rates of success or retention during a semester if they are placed in a course where they can reasonably attempt assignments and understand course material. This theory appears to underlie the belief in the utility of creating homogenous student groupings on the basis of scores and referring them to alternate treatments. The use of score-based tests to refer a proportion of a group to some treatment implies that it is effective primarily for those students while some other treatment is more effective for the rest. Referring a proportion of students to a remedial reading or language program while another group is referred a higher level implies an interaction between the type of treatment and the type of student. However, decision theory, while recognizing the importance of differing treatments for different abilities did not explicitly recognize the differing treatments that may exist within a level of treatment. For example, not all remedial classes at the same level are taught or structured the same way. Some courses may be taught by full-time faculty, others by part-time, adjunct faculty, and others may be computer laboratories that use tutoring software on individual computer terminals. Decision theory also did not explicitly recognize that individual aptitude, and other characteristics of the student, may interact in different ways in different classes. Differing characteristics of students, even after pre-sorting them by test score, can lead to different outcomes even though the instructional content may be similar.

Predictive validity is essentially an attempt to approximate the future in the present. Therefore, the greater the correspondence between the skills measured on a test and the skills needed for success in a course, the evidence of predictive validity is enhanced . -For example, Identical Elements Theory (Thorndike, E.L. and Woodworth, 1901) suggests that transfer in performance between a predictor (or prior task) is enhanced to the extent that

the predictor or tasks share common elements. According to Identical Elements theory, the ability to transfer between training and application contexts is the central element in determining the usefulness or application of a program designed to predict performance in the actual task (Kaplan and Saccuzzo, 1993). Knowledge emphasized in training or testing may not adequately simulate the tasks demanded later in school or on the job. The differences may involve situational characteristics external to the task. Differences might include social interaction patterns, stress or ambient noise, explicit or implicit rules that govern task performance, the range of variation in the stimulus environment, the nature of available responses, performance schedules, or characteristics of the performer such as motivation, fatigue, or stress (Thorndike, R.L. 1991).

Identical Elements Theory suggests that a major technical challenge facing educational and training practitioners is the need for similarity between the testing, training, and performance contexts. If the context exactly simulates the performance context, transfer of knowledge and skills should be perfect. Therefore, the more shared elements, then the more similar the two tasks, hence the more transfer there would be. For example, research in personnel selection has shown that paper and pencil selection and classification tests have much higher predictive validity when the criterion is a test of job knowledge than when the criterion is supervisor ratings (Jensen, 1993); Stitch, 1975; Asher & Sciarrino, 1974). Although it is unclear how Thorndike defined identical elements, some scholars believe that his theory refers to mental elements that are transferable from one situation to another (Druckman and Bjork, 1994). For the purposes of understanding predictive validity and to interpret evidence derived from studies, Identical Elements Theory helps to provide a interpretive framework.

Pertinent to this investigation and to an understanding of the relationship between predictor and criterion is Point-to-Point Theory. As described by Asher and Sciarrino,

(1974) this theory suggests that the use of cut scores on a placement test to predict the often normative outcome variable of final grade will generally result in unreliable coefficients. As stated by Asher and Sciarrino (1974), "Information with the highest validity seems to have a point-to-point correspondence with the criterion " (p. 519). In their analyses, Asher and Sciarrino demonstrated that work sample tests that were in effect miniature replicas of the actual job or task for which prediction of aptitude was desired consistently demonstrated higher validity coefficients. This work demonstrated strong support for the use of work samples for selecting employees in a variety of areas, including motor skills.

Knowledge of how the student or applicant performed in similar settings and how the criterion for success was determined greatly improves the validity of selection and placement decisions. This should be particularly true according to Asher if there is correspondence between the predictor and criterion space. For example, historical information about job applicants was the best predictor of future performance in specific positions. Factual and verifiable biographical information showed substantially higher predictive validity when job proficiency was the criterion, than other predictors such as intelligence tests, personality, interest, motor skill and mechanical ability inventory tests. These observations and theoretical considerations were of particular value in the formulation of this investigation.

Aptitude-, and Trait-Treatment Interaction Theory (ATI or TTI)

Cronbach and Snow (1969) extended Decision Theory to focus more closely on variations in instructional methods and the interaction of this variation with student characteristics. As in decision theory, aptitude-treatment interaction theory focuses on the processes and outcomes associated with student referral to alternate educational treatments. However, as originally described by Cronbach and Snow, aptitude-treatment interaction

focuses on the importance of evaluating these alternate educational treatments. Under this theory, questions of the effectiveness of a placement system focus on how the characteristics of the treatment such as a remedial reading program, or an honors program interact with the characteristics of the individual student. This observation greatly informed the design of this investigation.

In higher education, Willingham (1974) discussed and applied aptitude-treatment interaction theory to placement within an instructional sequence. Willingham noted that the interaction of individual traits, characteristics, and abilities, with particular forms of treatment is pivotal to successful institutional adaptation to individual differences. In applying ATI to the analysis of placement decisions and predictive validity in higher education, Willingham emphasized the identification of how students were grouped, the variation in treatment, and the assessment of learning outcomes. This study will adopt this approach in developing and enhancing a model for student placement in the community college.

ATI posits that individuals interact and respond to treatments differently. Some types of individuals respond favorably to one treatment while other types respond favorably to another treatment (Willingham, 1974; Snow, 1972). Each referral to a particular treatment based on a test score or placement rule represents a decision designed to enhance the individual or the institution or both. If a student is not eligible or allowed into a particular treatment, then he or she is directed into another course or "treatment." Thus the institution benefits from a more efficient use of teaching resources. In short, the usefulness of an educational placement system necessarily involves a comparative evaluation between alternate treatments.

In the years following the application of aptitude-treatment interaction theory to predictive validity studies, subsequent research led to the concept of a differential validity.

Differential validity was identified and applied to predictive validity studies in education by Thorndike (1971). Both ATI and differential validity stress the importance of interaction effects, and that the validation of this interaction must recognize differences between treatments. As noted by Cronbach (1971), "A 'validity coefficient' indicating that test X predicts success within a treatment tells nothing about its usefulness for placement" (p. 500). Although two sections of a particular class or training session may have similar objectives, the method, instructional media, approach, and particular characteristics of the instructor or setting may produce different outcomes. This may occur despite attempts at achieving homogenous student groupings through a placement and exemption system. Differential validity suggests that interaction of traits and treatment is central to judging the predictive validity of a placement system. As stated by Thorndike (1971): "As awareness of the requirements for the placement and classification use of tests in education increases, a different criterion will be applied to tests for these purposes - the criterion of differential validity."

Taken together, aptitude-treatment interaction theory and differential validity theoretically suggest that when student assessment is conducted in a thorough and systematic way, then educational decisions regarding student placement in the curriculum will serve to improve student outcomes. Improvement in outcomes will occur because the particular traits, skills, and demands reflected on tests should reflect a student's ability to perform similar tasks in a classroom, employment, or training setting. ATI and differential validity theory suggests that because the placement tests are normative, then the test scores can be used in a prescriptive way to assign students scoring in a certain range to different treatments deemed more suitable for their level of cognitive ability. Placement by ability is thus supposed to enhance the interaction effects of individual student characteristics with a particular level of the instructional sequence.

However, in practice, there are several confounds to this assumption. An assumption of classical test , decision , and aptitude-treatment interaction theories is that the criterion for judging success is stable and reliable. This assumption of criterion reliability is often violated. Grading and evaluation differences among faculty teaching the same level course confound this assumption (Lewis, 1978). Popham also noted that there exist several systems for grading. These include relative grading where students are compared to their peers in the class, absolute grading where attainment of a criterion determines the grade, comparing performance to aptitude, and student gains in achievement over time.

Another assumption is that students can be reliably grouped using a measure that sufficiently accounts for individual differences. This assumption of creating homogenous groups using score ranges is also often violated. This is because most commercially available standardized tests sample only a small portion of what a person knows or is able to do (Gifford, 1989).

An assumption of decision and aptitude treatment interaction theories is that the instructional method is appropriate for each student in a particular group. However, in the community college setting, there are reasons to suspect that the criterion used to judge validity may be unreliable, and that the information used to derive inferences about and place students may be under- or mis-specified. Moreover, there are reasons to suspect that within the same level of the curriculum, instructional treatments, approaches, and evaluation of student performance can vary. This study will attempt to address these confounds to predictive validity that may exist in the community college setting. This investigation will thus propose to extend the application of aptitude treatment interaction theory and differential validity to include variation within instructional level in determining student grades and retention.

Predictive Validity: The Need for Standards

This investigation relies primarily on the terminological and methodological conventions recommended by the American Psychological Association (APA), American Education Research Association (AERA), and the National Council on Measurement in Education (NCME). In 1954 the APA issued a publication entitled Technical Recommendations for Psychological Tests and Diagnostic Techniques. One year later the AERA and NCME published Technical Recommendations for Achievement Tests. These three professional organizations have since collaborated on joint publication and subsequent revision of the Standards for Educational and Psychological Testing (APA, 1985).

The *Standards* are particularly useful for this investigation because the decisions made about an individual's eligibility for a class or program of study are considered "high stakes" decisions. That is, the inferences made about an individual from a test score may influence their economic, social, or psychological well-being. For example, a person denied entrance into an occupational program on the basis of a minimum cutting score on a test may be negatively affected in their career choices and potential lifetime earnings. This was noted by Beusse (1974), in his study of the improvements in lifetime earnings made by low aptitude individuals allowed into the military services when cutting scores on entry tests were temporarily lowered. Low income individuals given the opportunity to enter the military forces during Project 100,000 (a War on Poverty program) were found to have higher indicators of socioeconomic improvement than a comparison group. The two groups (veteran and non-veteran) were compared on income, years of education, occupation and employment status. Another example of a potentially high stakes decision involving testing and access is federal Ability to Benefit (ATB) legislation. ATB regulations forbid students from receiving federal financial aid if they do not achieve

minimum scores on a standardized test. These students are essentially deemed unable to "profit from instruction."

The definition given by the *Standards* is a common definition used to describe the construct of validity - that is, the extent to which a test measures what it is intended to measure. However, many scholars in the field of educational measurement emphasize that validity is not solely a characteristic of a test but is a *function of the use to which scores from the test are put.* Cronbach (1971, p. 447) states:

> The phrase validation of a test is a source of much misunderstanding. One validates, not a test, but an interpretation of data arising from a specified procedure. A single instrument is used in many different ways--Smith's reading test may be used to screen applicants for professional training, to plan remedial instruction in reading, to measure the effectiveness of an instructional program, et. Since each application is based on a different interpretation, the evidence that justifies one application may have little relevance to the next. Because every interpretation has its own degree of validity, one can never reach the simple conclusion that a particular test *is valid.*

Popham (1990, p. 94) states it succinctly by observing that "a test is merely a tool that we use in order to make inferences. If we misuse the tool by employing it in the wrong situation, it is not the tool that defective. It's the tool user." Linn (1977) echoes this assertion when he states that "questions of validity are questions of the soundness of the interpretation of a measure.". Thus the concept of validity in educational testing is concerned with the validity and *defensibility* of the score-based inferences we make from the test score.

The potentially high stakes nature of testing also has implications for the present investigation. In the community colleges and higher education in general, an individual classified as a failure or low-aptitude and placed in a non-credit, basic skills course in

college, may have opportunity costs associated with prolonging their college career. These opportunity costs may be further aggravated by ineligibility for financial aid. A student could also be barred from entering an occupational program based on low test scores. The classification of the student as unlikely to succeed may negatively affect their future career choices and potential lifetime earning. Thus the high-stakes nature of assessing aptitude, predicting future performance, and classifying individuals on their likelihood of success in school militates for the use of recognized and agreed-upon standards for validity evidence.

The usefulness of the *Standards* in evaluating testing, admission, and placement practices is increasingly recognized outside of the educational community. According to Popham (1990), because of the likelihood that the use of certain tests will be challenged in court, many test developers and users are relying on the *Standards* to avoid appearing inconsistent with professional test evaluation guidelines. For example, as described in Chapter One of this dissertation, there have been many legal challenges to the use of tests to allocate opportunity or advancement in the workplace. In courtroom challenges, the adequacy of educational tests in allocating opportunity is increasingly judged by references to the *Standards*. The State Chancellor's Office of the California Community Colleges has also adopted the *Standards* as the basis for judging the adequacy of college testing and placement programs. The *Standards* help to provide a framework for gathering and evaluating validity evidence and are used as the framework for guiding the inquiry into predictive validity for this investigation.

Predictive Validity in the Community Colleges: Some Recent Studies

Over the last several years, there have been many studies that have attempted to establish the predictive validity of test scores in the community colleges. With the advent of the matriculation legislation in California for example, many of the 107 community

colleges have had to produce evidence of the predictive validity of placement tests and practices (Alkin and Freeman, 1992). However most of these studies have found the validity coefficient to be low to moderate. Although several studies were reviewed for this investigation, selected predictive validity studies are described below as illustrative of the general findings of most of these studies.

Many prior studies have focused on the criterion-related predictive validity of placement tests as they relate to course grades, however, many of these same studies have found small or modest relationships between final grade and placement test score (Hills, 1971; Cohen, 1987; Bridgemen, and Wendler, 1989; Gabe, 1989; Hodges, 1990; Hughes and Nelson, 1991; Isonio, 1991 and 1992; Hudson, McPhee, and Petrosko, 1993; Rasor and Barr, 1993, College of the Canyons, 1994; Armstrong, 1995). There are studies that focused on the effect of having a prerequisite on course success in post-secondary educational settings. (Iadevaia, 1989). With respect to the effect of having a prerequisite on course outcomes, the findings from these studies are generally mixed. Some show a statistically significant relationship between having the prerequisite and course grade, but the practical differences in success rates are modest.

At College of the Canyons in California, researchers found the median correlation coefficient between scores on the College Board Assessment and Placement Services (APS) Writing test and final course grades to be approximately .34. However, the coefficient for the APS Reading test and final grade was .14. The coefficients reported in their study had been adjusted for restriction of range. These findings suggest that the APS Writing test is a low to moderate predictor of final grade in selected English courses, while the Reading test predicts poorly. Both scores were below the .35 coefficient threshold required by state educational code. (College of the Canyons, 1994).

In the early 1990's Isonio (1991, 1992, 1993) conducted predictive validity studies for various English, ESL, and mathematics placement tests used for placement at Golden West College in California. These studies also focused on the predictive validity of the APS Reading and Writing, and Mathematics Diagnostic Testing Program (MDTP) in relation to student final grade in selected English and mathematics courses. With respect to the predictive validity of the APS tests to grades in English courses, Isonio's evidence followed a pattern similar to the College of the Canyons' studies cited above. The APS Writing test showed a low to moderate correlation coefficient, while the APS Reading test was a low predictor. With respect to mathematics placement, the MDTP tests were found to have a correlation coefficient of .37. This coefficient exceeds the state required standard of .35, although it is still a moderate-sized coefficient. Little difference was found in the coefficient between test score and outcome when a measure of retention (the withdrawal grade notation) was included in the criterion variable of GPA. Using instructor evaluation of the accuracy of student placement rather than final grade as the criterion, Isonio found relatively stronger predictive validity evidence for the ESL placement tests used to predict performance in ESL courses.

Dispositional data were also collected and used as a predictor in the mathematics study conducted by Isonio. He found that dispositional data gathered with a questionnaire such as grade expected in a course, or grade received in their last mathematics course, had approximately the same correlation with final grade as did test scores. This finding supported observations by Sheldon (1970) almost twenty five years earlier. Sheldon concluded that the predictive power of standardized tests was little better than the high school GPA of the student in explaining variance in course performance. This is perhaps further evidence of the value of biographical variables that share a point -to- point correspondence with the criterion.

In their study of the predictive validity of ASSET scores for student course outcomes at Riverside Community College in California, Hughes and Nelson (1991) found low predictive validity coefficients. They analyzed student scores from ASSET tests in reading, writing, and mathematics. They found a correlation of .23 between outcomes in English composition and ASSET writing test scores. The correlation between success in freshman composition courses and the ASSET reading test was actually negative (-.09). Hughes and Nelson recommended that ASSET test scores not be used as predictors for student success in the community college. They further recommended that additional dispositional data about the student be used in the placement decision given the low observed association between test scores and course performance outcomes. These dispositional data might include previous school performance such as high school or college course grades.

Predictive validity studies using different tests conducted at colleges outside of California show similar patterns. Of the several predictive validity studies reviewed for this investigation, the study conducted at Terra Community College is an example of the pattern and general findings of these various studies (Gamble, 1994). Researchers analyzed the relation between scores on American College Testing (ACT) ASSET test and cumulative student grade point average. Using a convenience sample of less than 100 full-time students, they found a significant correlation between various subtests of the ASSET test battery and cumulative GPA. They concluded that although the correlation coefficient was statistically significant, the findings could not be considered conclusive. The amount of common variance between test scores and cumulative GPA was in the range of 16%. No evaluation of the predictive validity of ASSET scores in relation to student retention was conducted, although such a study had been recommended by the authors.

Gabe (1989) also attempted to empirically demonstrate the relation between college course performance and scores on the ASSET placement test. Gabe's study found that the ASSET test scores predicted no better than student self-assessment of ability in college level writing. This finding of the relation between dispositional predictors and course performance supported the observations of Isonio cited above.

Gabe (ibid.) also examined the performance of students who attempted courses without what were considered necessary prerequisite courses. She found that students who entered a college level course without first taking the necessary preparatory course in writing passed at only a slightly lower rate than those who took the course. In one instance she noted that students who enrolled in a particular English course without the preparatory writing course succeeded at a higher rate than students who completed the prerequisite. In most cases, the success rates of students without the prerequisite was almost the same as those with the prerequisite. On the basis of this evidence, Gabe recommended a "flexible" placement approach which included information pertaining to student self assessment in addition to other proxies of ability in the placement decision. In her words, "mandatory placement into, or out of, college preparatory classes based on passing standards with questionable predictivity may not be the ideal solution for students." (p. 6).

In a course placement study reported by Jenkins (1991), scores on the ASSET test were compared with course grades at Redlands Community College in Texas. The observed association between ASSET scores and course outcomes in freshman composition was very low. This suggested to Jenkins that the ASSET may not be a good predictor of course performance.

The general findings from the studies cited above are illustrative of the trends found in other studies that reported the predictive validity of standardized tests in relation to course outcomes. Although many studies mention the importance of course retention, few

of the studies reviewed for this investigation used retention as a criterion variable for predictive validity.

Why do predictive validity studies in community college settings typically yield low predictive validity coefficients? Part of the answer may lie in the criterion used to judge the predictive validity of placement testing. Confounds in the criterion of final grade or retention are not a new issue in education. Low associations between scores on tests and success in college courses has been observed as a problem for many years (Hills, 1971, Cohen, 1987). The problem of predicting success is not limited to education. A similar problem confounds the criterion-related validity evidence of tests used for personnel selection, classification, training, and attrition in a variety of public and private settings. There have been many studies done in employee selection and training that confirm the difficulty of predicting on-the-job criteria and success in training (Sticht, 1992; Sticht, 1975; Asher, 1974; Cronbach and Gleser, 1965).

Similar findings were found by Sticht (1975) where he and associates identified the use of Skills Qualification Tests (SQT) for use in the military services over paper and pencil tests. The SQT was better in his view because the test replicated a sample of the tasks required on the job. His assertion was supported by evidence showing higher correlations with performance on the actual job for which the SQT was designed to predict. These findings provide support for the viability of Point-to-Point theory in predictive validity studies.

Predictive Validity and Predicting Future Behavior: Past as Prologue

With respect to this investigation, specifically Hypotheses 3 and 4, it has been noted that certain verifiable biographical information shows a significant positive relationship with education or training outcomes (Asher, 1974). In his review, Asher cited

research suggesting that over the last half century, the best single predictor of college grade point average is high school grade point average. This is consistent with Asher's contention that information with the greatest validity shares common features with a criterion, or in his words, a "point-to-point" correspondence. Supporting research can be found in studies that suggest a strong correlation of high school GPA with college GPA (Holland and Astin, 1962; Astin (1975; 1992).

Holland and Nichols (1964) also found a point-to-point relationship between what the student did in school and his or her subsequent behavior in college. For example, beyond predicting college grades, they found that scientific activity (field trips, participation in special science clubs or organizations, school sponsored activities in the sciences.) in high school predicted similar behavior in college. These findings on activities and behaviors in high school also held for involvement on campus, student leadership and participation in the fine arts in college.

With respect to explaining variance in course retention, theory was also instrumental in positing relationships among the independent and dependent variables used in the present investigation. After review of other student retention models (Bean, 1982; Pascarella and Terenzini, 1991; and Tinto, 1993), the dispositional and situational categories used by Cross (1985) appeared to be the most useful for this investigation.

Other models of student retention used assumptions about the student population that more generally describe four year college students rather than community college students. In her analysis of factors related to student retention and persistence, particularly for adult students, Cross identified three primary categories of variables. On the individual level, she found that adult student retention was related to two groups of variables that were defined as the *dispositional* and *situational* characteristics of students. Cross identified a third group of variables as *institutional* reasons for student departure.

The participant sample in this study more closely resemble the characteristics of the study populations examined by Cross. For example, many of the student level variables identified by Cross focus on the characteristics of adults returning to education after an absence. Many of the students in her studies were non-traditional, part-time, lower income students, with uncertain goals in school. Many were returning to school after a prolonged absence and had family and job responsibilities. Review of the demographic characteristics of the participant sample for this study suggests that the populations used by Cross appear to more closely fit the student population attending the four community colleges included in this study.

Institutional Factors

According to Cross, institutional reasons for student outcomes include the practices and procedures that may discourage or exclude adult students from attending school. Institutional reasons might include, inconvenient scheduling of necessary courses or location of classes, maintaining a full course load or progress requirements, and mandatory prerequisite courses or skills required before enrollment in a course or program of study. Other researchers have identified institutional characteristics that may affect student outcomes in higher education. Institutional variation also can include the prestige of the institution, size, public or private control, selectivity, proportion of students living on campus, institutional mission, and affiliation (Lenning, 1982). This study does not employ measures of institutional variation because it is using subjects from a single type of institution; a public, open-access two year community college. Thus there is little variation on these traditional measures of institutional effects. It is here that the Cross approach is also more appropriate for this study. Her definitions of institutional reasons for student outcomes are based more on the practices and policies of the institution. In this study,

variation in institutional policies and procedures might best be represented by a change from advisory to mandatory course placement and grouping of students for instructional purposes. It is therefore possible to examine the potential effect of a policy where students who are predicted to fail in fact have the opportunity to attempt a course. The effect of this institutional policy can therefore be more closely examined.

Situational Characteristics

According to Cross, a situational factor or reason for adult student performance in school are those that arise from a student's life situation at a given time. Situational reasons for attrition or course performance are often unpredictable and frequently result in a lack of time to devote oneself adequately to educational pursuits. Situational demands may include family needs such as child care, the needs of a spouse, changes in work schedules or job responsibilities, transportation problems, moving, or medical problems. Lack of money is a problem for young adults and others of low income and they often must work part or full-time to earn money to support themselves or their families. This situation frequently confronts community college students (Breneman, 1981).

One example of a situational indicator of a student with family or job responsibilities might be the time of day they attend classes. Many community college adult students who work or have family responsibilities can only attend in the late afternoon or evening. Thus students attending at these times may have family or job demands that affect their performance or retention in college. Knowledge of the time of day students attend classes may be an insight into the situational demands of the part-time, working adult student. Students who attend part-time, attend on a non-continuous basis, and must often fit educational pursuits into their increasingly complex lives increasingly the rule rather than the exception in community colleges (Warren, 1985).

The situational characteristics of students may also be inferred from student interest or need for support services as indicated on a questionnaire. The questionnaire used in this investigation contains items that ask subjects to indicate which college support services they are interested in receiving information about. An indication of interest in certain support services could be inferred as an indicator of a situational need of the student. For example, a student experiencing situational difficulties related to lack of money may receive or indicate interest in financial aid or job placement services. Those with family responsibilities may indicate an interest in information about child care. Physically or learning disabled students may indicate a need for services for disabled students. In these cases an expression of interest in a student support service such as financial aid or job placement was placed into the situational category of variables and entered as a block into the model explaining course grade and retention.

The support network of friends or family may be an important variable in whether a student persists in school. Astin (1994) found that student performance in college was strongly affected by peer group influences and support for the student. In her review, Cross (ibid.) also noted that for many non-traditional adult students, a supportive network of family or friends is a situational factor that often contributes to student success.

Dispositional Characteristics

Dispositional variables that may be related to student retention or performance are those that come from the behaviors, attitudes, self-perceptions, and abilities of the student. According to Cross (1981), adult students returning to school after a prolonged absence, or those who did not do well in school before may lack interest in learning or confidence in their ability to learn. Many community college students may be uncertain of their ability to perform adequately in school. Cross's identification of dispositional factors appears to be

highly appropriate as a method to organize and model the effect of this category of variables on student course outcomes. This may be important if there is merit to the theory that present behavior is best approximated by indicators of past behavior in a point-to-point correspondence.

Insight into the dispositional characteristics of the student may be found in some of the prior activities and behaviors of the student. Students who chose to complete high school with a diploma may have certain dispositional traits that enable them to persist in school. Research conducted in the military services suggests that first year attrition and performance is strongly related to the high school diploma status of the recruit. Recruits with high school diplomas may earn lower scores on entry exams and still have preference over higher scoring non-diploma graduates who possess a GED or high school proficiency certificate (Eitleberg, 1984). This is because actuarial data suggest that high school diploma graduates have lower attrition rates and higher rates of first year persistence than non-diploma graduates (Sticht, et. al., 1987). This may be due to the attitudinal or dispositional characteristics of the high school completer. They have demonstrated the ability to cope with organizational demands, schedules, and rules. The bright, but erstwhile non-diploma high school equivalent graduate may have higher mental ability or test scores, but was unable or unwilling to accede to the demands of attending class or completing assignments on time. The method of high school completion may provide insights into the disposition of the individual.

In California, students may earn a high school diploma through attaining a passing score on the California High School Proficiency Examination. (CHSPE). Earning passing scores on a high school equivalency exam however may not be a reliable predictor of how students will perform once they return to school. A student may achieve high scores on standardized tests, but lack other dispositional characteristics important to success in school

such as motivation or accommodation to rules of attendance. To use the theoretical rationale developed in a prior chapter, behavior associated with leaving high school early may have a greater Point-to-Point theoretical correspondence with retention behavior in the community college. This even though the student has demonstrated the cognitive skills at or above that of a high school graduate as measured by scores on equivalency exams.

Another example of a dispositional trait may be found in a the student's veteran status. Students who entered the armed forces and earned the status of a veteran may have traits that favorably predispose them to complete a college course. With its emphasis on rules, order and discipline, completion of a military term of service may have a greater point-to-point correspondence with student propensity to complete a college course. Veterans status may also be an indicator of the age and maturity of the student. With greater maturity and perhaps financially buoyed by GI benefits, students may have additional motivation and resources to persist and do well in their educational pursuits.

Knowledge of student educational goal or intent may be useful as a dispositional indicator of the student. Students attending for personal enrichment or avocational reasons may have differing levels of commitment or motivation to succeed or stay in class. Knowledge of student educational intent may provide useful information into improving the prediction of course outcomes.

Students returning to school after an absence may have negative beliefs about their own ability to succeed in school. Many adults re-entering school may feel that they are too old to learn or begin a new program of study that may take several years to complete. Also, adults with poor educational backgrounds may lack interest in learning or confidence in their ability to learn.

According to Cross, student dispositional characteristic include the prior preparation and academic abilities of the student. Indicators of the past performance of the study

participants in school may be helpful in predicting or explaining performance in the future. This is also consistent with Asher and Sciarrino's contention that future performance is best understood in the context of similar performance in the past. An example of how Point-to-Point Theory and dispositional predictors of performance can inform student placement is knowledge of past performance in other school settings. For this investigation, participants are asked to provide their estimated high school grade point average, and the grades they earned in their last English or mathematics course. It was also observed by Sheldon (1970), that for community college placement, predictive validity may be made optimal through the use of high school grade point average rather than scores on standardized tests.

Although high school GPA might best be viewed as a proximal indicator of student preparation, a more distal indicator of preparation might be the high school attended by the study participant. Freeman (1976) observed over 20 years ago that the high school diploma was losing credibility as an indicator of achievement or aptitude. Much the same criticism was leveled at secondary education in the report, A Nation at Risk (National Commission, 1983). Researchers have also noted that not all high schools produce similar outcomes (Oakes, 1985).

Recently the San Diego Unified School District (SDUSD) ranked its high schools as part of a new accountability system (Jones, 1997). Schools were ranked based on average school scores on an English-language standardized test and grades assigned by teachers. High schools were also ranked by the proportion of students taking college preparatory courses. The schools receiving the lowest ranking were placed under a "performance review," (also called PAR in their study) and required to take steps to improve their academic performance. To gather additional dispositional and academic background data for this study, the students in this investigation from the 19 schools were divided into two groups. One group was comprised of the lowest ranking schools that

received performance review status (PAR schools). The other group included the higher achieving schools that were not placed under a performance review (Non-PAR). In the PAR group there were five high schools identified, while the non-PAR group included the remaining 14 high schools. The five lowest performing high schools in the SDUSD are primary "feeder" schools that transfers many students to the community colleges used in this investigation. This grouping was intended to serve as a distal indicator of dispositional or academic preparation data for the students included in this investigation. Students from the PAR schools were expected to perform below the students from non-PAR schools. The distal indicator of high school preparation was included in the course performance model as a dispositional indicator.

Summary

A review of predictive validity studies conducted in two year colleges shows a general pattern of relatively low validity coefficients and inconclusive and often counter-intuitive findings. Few of the studies reviewed included retention as a course outcome, even though many college implemented placement system in response to faculty complaints about high attrition in composition courses due to inadequate student preparation. To attempt to fill this gap in the literature, this study tested the relation between test score and student retention. Several of the studies also recommend the use of certain biographical variables for use in the placement decision, although few present findings to support this recommendation.

The work of Asher in support of a Point-to-Point theory and Cross in identifying traits or behaviors of the student suggest that student biographical and trait information has potential value in student placement. Point-to-Point theory was used as the foundation for identifying and gathering relevant student data. Cross's taxonomy of student dispositional

and situational variables was used to make this theory operational in this investigation. The Cross model was of particular value in parts two and three of this investigation.

In part IV of the investigation in chapter 5, using the theoretical framework of the interaction of student traits and instructional treatment suggested by Aptitude Treatment Interaction and the biographical and trait information suggested by Point to Point theory, the present constructed a model to explain variance in student course grade and retention. This model was intended to account for instructional and grading variation by including the instructor as a source of variance in predicting student performance. The use of ATI attempts to quantify the extent of grading variation and how predictive models for student placement might be made more reliable through knowledge of instructor variance in evaluating the performance or contributing to the retention of students.

CHAPTER FOUR

METHODOLOGY

Overview of the Research Design

This chapter describes the method and stages of the analysis designed to gather evidence for each of the research questions. The hypotheses derived from the research questions presented in this study posit several important interactions between the independent variables of placement test score, student biographical characteristics, instructor characteristics, and the dependent variables of course grade and course retention. The inferences made about individual aptitude and ability on the basis of scores and biographical data were tested empirically using two dependent variables; course grade and retention.

The investigation was conducted in four parts. Part I of the study focused on descriptive data about the participant sample. Demographic data are given for participant samples in both English and mathematics. Measures of central tendency are given for the dependent variables used in this investigation.

Part II of the investigation examined the validity of standardized placement test scores in predicting student course outcomes as measured by course grade and retention. Hypothesis 1 tested if test scores were statistically related to grades and retention in college English and mathematics courses. To gather evidence for Hypothesis 1, student scores on standardized placement tests were compared with their course grades in community college English and mathematics courses. Predictive validity coefficients were obtained from analyzing the correlation between test scores and final course grades. The correlation coefficients were analyzed by three levels of English and mathematics within the

curriculum. The coefficients obtained in specific levels of English and mathematics were corrected for restriction of range due to the potential problem of truncated range in test scores. The corrected coefficients were used to either accept or reject null Hypothesis 1.

Hypothesis 1a posited that the predictive validity coefficients would fall below the .35 level required by state educational statutes. When analyzed by curricular level, this part of the analysis also used the corrected coefficients to either accept or reject null Hypothesis 1a.

Part II of the analysis also included a testing of the point-to-point correspondence of test scores. This was the focus of null Hypothesis 1b which stated that the coefficients between successive administrations of a standardized placement test will not show a stronger relationship than the coefficient between test score and final grade. This hypothesis was made operational by two separate administrations of a standardized placement test to a group of students. The resulting pairs of scores were compared to determine the validity of one placement test score in predicting the second placement test score. The first administration occurred at the start of the fall semester and the second at the start of the next spring semester. Comparing these results for 92 pairs of scores was intended to provide evidence to either accept or reject the construct of a point-to-point correspondence between the two scores. The resulting coefficient between test scores was compared with the coefficients derived from the correlation between test score and final grade to provide evidence for either accepting or rejecting null Hypothesis 1b..

Part II also included the analysis of the effects of student eligibility status on course outcomes. This was the focus of null Hypothesis 2 that posited no relationship between course eligibility status and success in the course. The course outcomes of both eligible and ineligible students who were allowed to attempt a course are compared to test the effect of prerequisite status on course success. This was done using a series of two-by-two

contingency tables comparing the successful versus non-successful outcomes of eligible versus non-eligible students in seven subject areas. Evidence for either accepting or rejecting the null hypothesis was based on the magnitude of the phi coefficient.

The null Hypothesis 2a posited that the effect of course eligibility status would not be small to modest. Although there was not an appropriate test of significance for this hypothesis, evidence for either accepting or rejecting Hypothesis 2a was based on the relative success of the non-eligible to the eligible students. The success rates of the non-eligible were calculated as a proportion of the success rates of the eligible students. If the non-eligible students were found to be at least 70% as successful as eligible students, this was deemed sufficient evidence to reject null Hypothesis 2a and accept the alternate hypothesis that the effect of course eligibility status was small to modest.

Part III of the investigation focused on the relation of student demographic, dispositional, and situational variables to the course outcomes of final grade and retention. Null Hypothesis 3 was tested in this part of the analysis. Biographical data about the participants including demographic, dispositional, and situational data were gathered using a questionnaire administered at the time of assessment. Student biographical data were crosstabulated with the course grade and retention outcomes of the student to determine the extent and degree of association with course grade and retention. The chi-square residuals (the difference between the expected and observed counts in each cell) derived from the crosstabulation tables were inspected to determine the direction of the association between the independent and dependent variables. The chi-square statistic was used to determine the statistical significance of the association between the independent and dependent variable. The crosstabulations presented in this chapter are intended to provide an overall description of the content of the questionnaire with respect to the individual items, and the relation of participant responses to the dependent variables.

Crosstabulations of the independent and dependent variables were constructed for both the English and mathematics participant samples. Although not all the biographical variables were consistently found to have a significant relationship with retention and course success in both subject areas, all biographical items were selected for inclusion in the two explanatory models for course grade and retention. The crosstabulation tables were instructive as to the strength and direction of the relationship between the independent and dependent variables. However, prior research suggested a relatively high degree of interactive and mediating effects of student biographical variables (Pascarella and Terenzini, 1991; Elmers and Pike 1997; Pike, 1995). When there are interrelationships among the independent variables, the literature suggests the use of multivariate models to sort out interrelationships (Simon and Burstein, 1985). Therefore all demographic, dispositional, and situational variables were included in the two course performance outcomes models. Linear regression was used to construct a model that best explained variance in final grade. Because retention was coded as a binary variable, logistic regression was used to develop a retention model from the student test score and biographical variables .

Hypothesis 4 examined the role of the instructor as a source of variance in explaining or predicting course outcomes. This is the focus of part IV of the analysis. The model developed in part IV partitions the variance contributed by the five blocks of variables (test score, demographic, dispositional, situational, and instructors) that make up the model. Instructors are coded as dummy variables and entered into the model in the final stage after the entry of test scores, demographic, dispositional, and situational variables. The amount of error contributed by instructor grading variation in estimating the predictive validity of college placement practices was thus calculated. The change in R-squared noted by introduction of instructors into the models was used as evidence to either accept or reject

null Hypothesis 4 that posits no relation between instructor characteristics and final grade or course retention.

The relationship of the employment status of the instructor to explaining variance in the dependent variables was the focus of null Hypothesis 4b. This hypothesis posits no relation between the employment status (full- or part-time) of the instructor and the amount of grading variation accounted for by each group of instructors. To gather evidence for this hypothesis, two separate regression models were developed. One used full-time instructors as the final block of variables to enter the model, while the second model used part-time instructors as the final block of variables. The relative proportion of variance explained (change in R-squared) by each group of instructors was used to either accept or reject null Hypothesis 4b.

Part I: Descriptive Data about the Participant Sample

The first part of the analysis presents the demographic characteristics of the participant sample. Separate distributions are presented for both the English and mathematics samples. This is intended to provide an overall picture of the demographic characteristics of the participants with respect to sex, race, primary language, age, and learning disability status. Part I also included descriptive data about the measures used in this investigation to test predictive validity. The means and standard deviations of the three standardized placement tests used in this study are given. Measures of central tendency for each of the three placement tests (Reading, Writing, and Mathematics) are also broken down by curriculum level. If there was evidence that test scores dropped in lower levels of the curriculum, this was taken as evidence of a truncation of range in test scores. This evidence led to the decision to apply a correction for restriction of range to the predictive validity coefficients. The correction for restriction of range elevates the magnitude of the

coefficient to the level it would have achieved had there been no truncation of range or pre-sorting on the basis of cutting scores on the placement tests.

Descriptive data are also included for the two measures of final grade (GPA1 and GPA2) and retention. Grades were computed into grade point averages for the two scales, while retention was shown as a proportion of those retained in the course compared to those who dropped out of the course before the end of the term.

Part II: Predictive Validity of Placement Tests and Prerequisites in the Community College

The investigation began with the initial, policy-based questions on the relationship between standardized test scores and student outcomes:

> Are placement tests highly predictive of course performance outcomes as measured by course grades or retention in the class? How statistically significant is the relationship between standardized test scores and student achievement? Does this relationship demonstrate the state mandated .35 correlation coefficient? Is there a greater correspondence or correlation coefficient between test scores than between test scores and course grade?

A second set of questions focused on the use of mandatory skill prerequisites for entering a particular course. The research questions asked:

> What is the effect of mandatory prerequisite courses (e.g., completion of college-level English course in order to be eligible for enrollment in Sociology 100, or completion of intermediate algebra before one can enroll in a photography course) on student performance in the subsequent course? How do students without the mandated prerequisite perform in comparison with those having the prerequisite?

Table 4.1 presents a summary of the alternate hypotheses, the independent and dependent variables, the source of the data, and the analytical approach used to gather and evaluate predictive validity evidence.

Table 4.1: Alternate Hypotheses, Variables, Source, and Analytical Approach For Part II of Investigation

Alternate Hypothesis	Variables	Source	Analysis
1. There is a relationship between course performance outcomes as measured by final grade or retention and scores on standardized placement tests in community college English and mathematics courses .	Independent: APS Reading and English Placement Test Scores MDTP Mathematics Placement Test Scores	Student Test Score Database, Transcript, and Instructor Grade Rosters	For the relation of GPA and test score: : Correlation For Retention: the model chi-square improvement in the -2 log likelihood from logistic regression analysis
1a. The predictive validity coefficient as measured by final course grade or retention will fall below the .35 correlation coefficient mandated by California state Matriculation regulations for test validation.	Dependent: Course Grade, (A=4, B=3, C=2, D=1, F=0) (GPA1) and, (A=4, B=3, C=2, D=1, F and W=0) (GPA2) Retention: (0=Not Retained), (1=Retained)	Student Test Score Database, Transcript, and Instructor Grade Rosters	For the relation of GPA and test score: : Correlation Coefficients corrected for restriction of range when analyzed by course level
1b. Scores on successive administrations of the placement test will show a greater point to point relationship (i.e., a higher coefficient) than will the coefficient between test score and final grade	Score from two administrations of the DTLS placement test to participants	DTLS Test Answer Sheets	For the relation between placement test scores: Correlation and t-test

Alternate Hypothesis	Variables	Source	Analysis
2. There is a relationship between course eligibility status (i.e., a student has either completed or not completed a prerequisite for a particular course) and success in the subsequent selected community college courses.	Independent: Student eligibility for courses based upon completion of a prerequisite or minimum test score Dependent: Course Success (A, B, C, Credit=Success)	Student Transcript Database, and Instructor Grade Rosters	Bi-Variate Contingency Tables Phi statistic
2a. The effect of prerequisite status on success in selected courses will be small to modest .	(D, F, No Credit=Non Success)		Relative success of Non-eligible compared to eligible participants

The independent variables for part II of the present investigation were the placement tests and course eligibility while the dependent variables were the computed grade point averages (GPA1 and GPA2) and retention. In the case of Alternate Hypothesis 1b, the independent variables were the scores from the first administration of a placement test, while the dependent variables were the scores from the second administration of an alternate version of the placement test to the same group of students.

Instrumentation

The instruments used to derive test scores and placement information about study participants were the Assessment and Placement Services for Community Colleges (APS) placement tests developed by The College Board. There were two tests used to distribute students among the three levels of English curriculum in the study colleges. These were the APS Reading Comprehension Form A placement test and the APS Conventions of Written English Form B placement test. Both tests were written by The College Board and are among the most widely used placement tests in the California Community Colleges.

The format and content of the placement tests were drawn from the Comparative Guidance Program (CGP) which was a testing program developed in the late 1960's at The College Board. The Writing placement test is the revised version of the CGP and the Reading test was introduced in 1984. The technical manual supplied with the tests provide construct-, criterion-and content-related validity and reliability evidence (The College Board, 1984).

In selecting the APS tests the English department faculty in the college included in this study reviewed several placement tests for content-related validity evidence. After review of several placement tests, the committee selected the APS tests. English faculty on the committee generally agreed that these tests most closely reflected the skills needed to succeed in English courses in the three colleges offering college English instruction.

Adequacy of English Placement Tests Used in this Investigation

According to the technical manual supplied by the College Board as background to the APS tests, (The College Board, 1984) English faculty assisted in "developing the tests and establishing their content validity." (ibid., p. 21). This assistance took the form of identifying the skills they thought necessary for success in English courses as taught in open-admissions, post-secondary institutions. These instructors also approved all items used in the tests. The technical manual also provided information on the items used to address the content domain in community college English courses.

In addition, the College Board sponsored validity studies that are summarized in the technical manual and their findings are reproduced in table 4.2 below.

Table 4.2: Correlations of Test Scores and Grades in English Courses

			Correlation	
Course	**Predictor**	**Median**	**Highest**	**Lowest**
English	Reading	.28	.51	-.01
	Writing	.32	.55	.03
	Reading and Writing	.35	.57	.04

Source: The College Board, 1984

The validity studies conducted by The College Board relied primarily on correlations between test scores achieved by examinees prior to beginning college coursework and grades earned at the end of the first or second semester. English correlations were obtained from 64 validity studies, each of which had at least 100 students. The criterion variable is the final grade earned in the English course.

A review of multiple sources of evidence suggested that the APS placement tests were adequate for the purposes of this study. This finding is based on the information provided in the APS technical manual, the selection of the APS test for inclusion on the state approved list of assessment instruments, and its selection by community college English faculty for use in student placement. There are several tests that make up the APS battery of placement tests for community college students. Two of these tests are used to group and classify students for placement in the colleges included in this study. These are the APS Reading and APS Writing tests.

APS Reading Test

The 25 minute Reading test contains 35 four part, multiple-choice questions based on 8 reading passages. These selections, which are mostly written by contemporary writers use content drawn from the natural sciences, social sciences, and contemporary life.

They vary in style and exemplify different types of writing: straightforward reporting, persuasive writing, and description. Questions are intended to measure the student's comprehension of main ideas and specific details and their ability to make inferences and extract the meaning of vocabulary in context.

The test developer states that the achievement scores are meant to differentiate among students who are "adequately prepared for a college's academic work and those who may need developmental work, and they are useful in placing students in appropriate English courses." (The College Board, 1985; p. 6)

APS Writing Test

The writing placement test contains 40 multiple-choice questions intended to measure a student's ability to perform the "kind of writing usually required of students in colleges." (ibid.) The test questions assess abilities to recognize errors in grammar, usage, and word choice. The instrument appears to place emphasis on sentence structure, and clarity of expression in ideas and thoughts. There is no writing sample included in the assessment, the test relies strictly on responses to four-part multiple-choice items.

According to The College Board, a student scoring high on this assessment has a high probability of being able to write correctly and effectively at the college level The College Board cited prior research (Godshalk, 1966; (cited in College Board, 1984)) that indicated that scores on multiple choice questions similar to those used in the writing test are highly correlated to scores received by students on essay tests based on high inter-rater reliability.

To test Alternate Hypothesis 1b (correlation of test scores), the reading comprehension and conventions of written English placement tests from the Descriptive Tests of Language Skills (DTLS) (The College Board, 1989) were used. These placement

tests parallel the APS tests in content and format. The DTLS was created by The College Board to replace the APS tests in community colleges. The DTLS reading comprehension test consists of 45 questions administered in 45 minutes. Like the APS reading test, the DTLS reading comprehension test contains individual questions, and sets of questions based on reading passages. The test is designed to measure how well the student can identify word or phrase meanings through the context that they are presented in. The test is also designed to assess the subject's understanding of the author's literal and interpretive meaning, and the author's assumptions, opinions, and tone. This is similar in content to the content of the APS reading placement test.

The DTLS conventions of written English placement test consists of 40 questions administered in 40 minutes. As the title of the test suggests, it is designed to determine how much the subject knows about using standard forms of written English. This is also similar to the APS writing test.

Relation of the APS and DTLS

The APS was used for placing the subjects included in this study in 1994, and the DTLS was used in this study to test Alternate Hypothesis 1b concerning a point-to-point correspondence between test scores as compared to test score and course grades. Although the DTLS and APS are different tests, a review of both tests suggests that both are similar in form and content. To test the relation between the APS and DTLS and thus to judge the adequacy of the DTLS for this part of the study, students tested with the APS tests were also administered the DTLS tests in English classes approximately two weeks after taking the APS tests. Although there were some differences in test conditions between the two administrations, this approach enabled the collection of almost 500 pairs of scores. It was therefore possible to obtain the correlation coefficient between the two sets of scores. As

shown in table 4.3 below, the scores on the tests show a strong relationship. Thus the DTLS was determined as adequate for testing Hypothesis 1b.

Table 4.3
Correlation of APS and DTLS Scores

	APS Reading	APS Writing	N
DTLS Reading	.71		492
(sig.)	(p<.001)		
DTLS Writing		.75	492
(sig.)		(p <.001)	

Mathematics Placement

For mathematics courses, the Mathematics Diagnostic Testing Program (MDTP) Algebra Readiness Form A test was used to predict performance in mathematics courses. These tests were developed by a joint committee of mathematicians from the University of California and California State University. As with the APS placement tests, this test is among the most widely used in the California community colleges for placement into mathematics courses. A brief description of the placement test is included below.

The MDTP test used for placement into pre-collegiate and college level mathematics courses is titled the Algebra Readiness Test. This is a 50 item, 45 minute timed test designed to assess student knowledge of mathematical concepts needed for success in college level algebra courses. Test questions focus on student knowledge of the operations and applications of integers, fractions, decimals, exponents, square roots, simple equations and literal symbols, and geometry and graphing.

The MDTP test was written by a workgroup supported by the two public university systems in California (UC/CSU). The test item writers included mathematics and science faculty from the universities, community colleges, and high schools. According to the test developer, the primary goal of the MDTP test is to assess topics that are essential for successful outcomes in subsequent mathematics courses.

Adequacy of MDTP Placement Test

The MDTP test is widely used for placement in college mathematics programs. According to the publisher of the MDTP, every campus of the University of California, and one-half of the CSU and California Community Colleges use one of the MDTP forms for placement into the mathematics curriculum. To the test developer, this is evidence of both face and content-related validity. Similar to the APS tests, the MDTP test developers also published a technical manual that contains information on test reliability, content, and standard errors of measurement. Table 4.4 below summarizes data found in the MDTP technical manual and suggests that the tests meet the .35 correlation coefficient criterion for use in the community colleges.

Table 4.4: Correlations of Algebra Readiness Test Scores and Grades by Testing Site for Mathematics Courses

		Correlation	
Course	**Location**	**Highest**	**Lowest**
Elementary Algebra	California High Schools	.49	.42
	Community Colleges	.46	.35
	California State University	.62	.10

Source: Mathematics Testing Diagnostic Project, 1991

The information in table 4.4 suggests that the MDTP placement test is adequate for the purposes of this investigation. This finding is based on the information provided in the technical manual, and the favorable review given the test by college mathematics faculty.

To test the validity of the placement tests across different levels of course difficulty, student grades and retention outcomes included data gathered from English and

mathematics courses across three levels of difficulty for the three colleges included in this investigation. For English courses this included community college students placed and enrolled in pre-collegiate level courses considered "basic skills" (two levels below college, non-credit courses), one level below transfer (community college courses typically offered only for credit toward the associate degree), and transfer (accepted as equivalent courses at the University of California or California State University). Table 4.5 below shows the placement tests, course levels, course titles, and the number of cases available for analysis at each of the three levels.

Table 4.5: Placement Tests, Course Level, Course Title and Number of Cases Used For College Placement In English and Mathematics

Placement Test	Course Levels	Course Title	N of Cases
APS Reading Comprehension	Two Levels below Transfer (non-degree applicable)	College Reading and Study Skills	779
	One Level below Transfer (Associate degree-applicable)	Introduction to College Reading	2,113
	Transfer Level	Freshman Composition	1,033
APS Conventions of Written English	Two Levels below Transfer (non-degree applicable)	College Writing and Study Skills	779
	One Level below Transfer (Associate degree-applicable)	Introduction to College Writing	2,113
	Transfer Level	Freshman Composition	1,033
Total N for English			3,925

Placement Test	Course Levels	Course Title	N of Cases
Algebra Readiness Placement Test	Two Levels Below Transfer (non-degree applicable)	Mathematics Preparation Courses (e.g. Pre-algebra)	1,418
	One Level Below Transfer (Associate degree-applicable)	College Mathematics and Elementary Algebra	1,114
	Transfer Level	College Algebra, Trigonometry, Statistics, and Calculus	1,187
Total N for Mathematics			3,719

Dependent Variables

In part II of the investigation, criterion-related validity evidence was based on the extent to which an examinee's score allowed for statistically significant inferences about the student's final grade or retention. The standard for determining the strength of the relation between placement test score and grade was the .35 correlation coefficient defined in California state education statutes.

The dependent variables for part II of the study were:

1. The final grade for the course coded numerically using a traditional five point grade point average scale (A=4, B=3, C=2, D=1, F=0) This variable was called GPA1.
2. The final grade for the course coded as in GPA1 above but also including a Withdrawal (W) notation as equivalent to zero. (A=4, B=3, C=2, D=1, F and W=0). This variable was called GPA2.
3. Retention was coded as a dichotomous variable. A student not remaining in the course after the third week until the end of the semester (Withdrawal) was counted

as not retained and received a score of '0.' Students earning a grade at the end of the term were considered retained and were coded as a '1.'

Databases

For part I of the study two separate databases were used. One database is from the student information system of the colleges included in this study. This database contains course enrollment history, and a transcript for each student showing all grades received and retention information. The second source of data is from a student assessment database which contains student scores on placement tests and recommended referrals based on these test scores to courses at three levels in the curriculum. For this study, the investigator aggregated the various sources of data electronically using the student identification code as the matching variable for merging different datasets.

Analysis

Testing Alternate Hypothesis 1 was done by observing the predictive validity coefficients obtained through Pearson Product-Moment correlations of test scores (reading, writing, and mathematics) with GPA1 and GPA2. Statistical significance was set at the .05 level.

Because the criterion variable of retention was dichotomous, logistic regression was used to test the relation of standardized placement tests to retention. To test the effect of test scores in relation to retention, the chi-square of the decrease in the -2 log-likelihood was used to determine the significance of the change. The statistical significance of the test score variables in predicting retention, statistical significance was set at the .05 level.

Accepting or rejecting null Hypothesis 1a was determined by observing the magnitude of the predictive validity coefficient between test score and final grade. When analyzed by course level, the corrected *r* was used to test Hypothesis 1a to account for

truncation of range due to the pre-sorting of students by placement level. If the predictive validity coefficient (if r fell below the .35 mandatory level, the alternate Hypothesis 1a was accepted.

Evidence to either accept or reject Alternate Hypothesis 1b was based on the correlation coefficient between the two test scores. To accept Alternate Hypothesis 1b, the correlation coefficient between the pairs of placement test scores had to exceed the correlation coefficient between test score and final grades.

Participants

For analyzing the predictive validity of English and mathematics course placement, assessment data from all first-time matriculating students were gathered from the fall, 1994 testing period. In the colleges included in this study the testing for the fall term began in May, 1994 and extended until September, 1994. Of the approximately 6,000 students who had APS reading and writing assessment scores, 3,925 enrolled in one of the three levels of English courses at one of the three community colleges included in this study in the fall, 1994 term.

Of the approximately 4,000 students with mathematics scores, 3,719 enrolled in one of the three curricular levels of mathematics course in the fall 1994 term. As with English students, mathematics students were also tracked from assessment testing into courses to determine the level they chose to enroll in.

The Point-to-Point Correspondence of Test Scores

Part II of the investigation included analyses that focused on the point-to-point correspondence of test scores compared to the correspondence between placement test scores and final grade. Alternate Hypothesis 1b posits that the coefficient between

placement test scores will be stronger than the coefficient between test scores and course grade. This was made operational in this investigation by two separate administrations of a standardized placement test to a group of participants. The first administration occurred at the start of the fall semester and the second at the start of the next spring semester. The correlation coefficient between the two pairs of scores was compared with the coefficient observed between test score and final grade. If the coefficient between the pairs of scores exceeded the coefficient observed between test scores and final grade, then Alternative Hypothesis 1b was accepted. In addition, a t-test on the difference in scores was performed to determine the significance of any observed change in scores after a semester of instruction.

The Relation of Course Eligibility to Course Outcomes

Alternate Hypothesis 2 posited an effect on course outcomes for students enrolling in a class without a mandated prerequisite course or level of skill deemed essential for success. The success rates for eligible students are compared to the success rates for ineligible students.

To test Alternate Hypothesis 2, students were divided into two groups for comparison. One group included students who had met the required prerequisite (eligible students), and the second group were students who had not satisfactorily completed the prerequisite (non-eligible students). The course performance outcome for the two groups was measured by final grade ('A', 'B', 'C', or 'Credit'=Success). Non-success was determined by earning a 'D', 'F' , 'No Credit', or 'W' (withdrawal late in the semester) grade notation. For this part of the investigation, retention was not analyzed separately, but was included as part of the course outcome variable of non-success . An example of a

two-by-two contingency table used to test Alternate Hypothesis 2 is shown below in table 4.6.

Table 4.6: Design of Two-by-Two Contingency Tables for Analyzing Effect of Course Prerequisite on Student Course Performance Outcomes

	Outcome		
Eligibility Status	Success (A,B,C, Credit)	Non-Success (D, F, W, No Credit)	Total
Eligible (met prerequisite)	Eligible and Successful (correctly placed)	Eligible and not-successful (false negatives)	Eligible
Not-Eligible (did not meet prerequisite)	Not Eligible and successful (false positives)	Not Eligible and not successful (correctly placed	Not Eligible

Computation of the phi statistic based on the distribution in the two-by-two table enabled the estimation of the effect of eligibility status on the student success outcome and provided evidence to either accept or reject the Null Hypothesis 2 of no relation between eligibility status and course performance outcomes.

Inspection of table 4.7 below indicates that that in some cases, the analysis of several semesters of data resulted in large sample sizes, particularly for English and mathematics. It is anticipated that because the phi statistics is sensitive to sample size, that trivial differences in success rates would result in a statistically significant relationship between eligibility status and course success. Alternate Hypothesis 2b posited that the actual differences in success rates between the two groups would be small to modest. Although no precise definition is given in the literature to determine if observed differences are significant in a practical sense, this study employed the success rate of the ineligible as a proportion of the success rate of the eligible students. Also, some community colleges in California have adopted a standard that states that students with the prerequisite are to succeed at twice the rate as the non-eligible students and that no more than one-third of the

non-eligible students should be successful in the course for a prerequisite to be valid (Chabot College, 1996; Isonio, 1995).

Table 4.7 shows the subject areas chosen for the analysis of the effect of a prerequisite on course outcomes conducted to test Alternate Hypothesis 2. The table also shows the required English or mathematics prerequisite for entry into the course. As indicated in the table, course eligibility could be met either through completion of a lower level course in English or mathematics, or through achieving a minimum cutting score on a placement test. Table 4.7 also includes the number of participants included in the sample for each subject area.

Table 4.7: Subject Areas, Participant Eligibility Status and Required Prerequisite Course or Test Score used to Determine Effect of Course Eligibility on Course Performance

Subject	Course Eligibility Requirement (Test Score or Course)	Number Eligible	Number Ineligible	Total
Accounting	English and Mathematics	2,081	160	2,241
Chemistry	English and Mathematics	1,530	298	1,828
Economics	English	3,279	963	4,242
Engineering	Mathematics	137	78	215
All English Courses	English	434,566	54,770	489,336
Pre-College Writing	English Writing	5,258	891	6,149
Pre-College Reading	Reading Comprehension	5,725	1,181	6,906
College English	English	10,834	3,987	14,821
History	English	3,310	567	3,877
Mathematics	Mathematics	101,403	30,640	123,043
Office Information Systems	English and Mathematics	272	117	389

For example, as shown in table 4.7, a student is declared "eligible" for a course if he or she has completed a course or achieved a minimum cutting score on a placement test in English or mathematics deemed essential for success in the subsequent course.

Ineligible students are those who have either not completed the required English or mathematics course to enter the course nor achieved a minimum cutting score on the relevant placement test. In the subject areas shown in the table, there were a large number of students who attempted courses without the desired prerequisite or skill level. Some of the courses in the analysis required an English skill level prerequisite, some a mathematics skill prerequisite, while some required both an English and mathematics prerequisite.

Accounting, Chemistry and Office Information Systems courses required both an English and mathematics skill level prerequisite. English, History, and Introductory Economics courses required only an English skill level prerequisite. The courses in Engineering required a mathematics skill level for students to eligible for the course.

Participants in Course Eligibility Analysis

To analyze the effect of eligibility status on course outcomes in the seven subject areas identified, a different group of study subjects was used. This part of the study used information from different subject areas and a student did not necessarily have to take a placement test nor enroll in an English or mathematics course to be included. For many subject areas outside of English or mathematics, students did not necessarily need a test score to meet a prerequisite for a course. This analysis also covered a longer time frame than a single semester. Data were aggregated for the eight semester time period during which students could more easily ignore placement advice and attempt a course that they were technically ineligible to enroll in.

Part III

Composite Model of Student Aptitude and Biographical Information for Predicting College Course Outcomes

In part III of the investigation, the study added student biographical data to the prediction of course performance to determine how much, if any, improvement could be made in the prediction of student achievement and retention. Student biographical variables related to course success and retention were grouped into demographic, dispositional, and situational categories. These variables, once grouped into these categories, were included in the explanatory models for course success as measured by course grade and retention.

Part III focused on the questions:

> How do student characteristics affect the prediction of performance outcomes? How much can predictive validity be improved by including student biographical information into a model that explains college course performance?

Table 4.8 presents a summary of the alternate hypotheses, the independent and dependent variables, the source of the data, and the analytical approach used to gather and evaluate evidence for Alternate Hypothesis 3 .

Table 4.8: Alternate Hypotheses, Variables, Source of Variables, and Analysis for Part III

Alternate Hypothesis	Variables	Source	Analysis
3. There is a relationship between student background and biographical characteristics and course performance outcomes as measured by course grade and retention after controlling for test score.	Independent: *Placement Test Scores*: APS and MDTP Tests Independent: *Demographic* Ethnicity, Gender, Age, Disability, Primary Language, Learning Disability,	Student Test Score and Demographic Databases .	Cross-tabulation, Correlation and

Alternate Hypothesis	Variables	Source	Analysis
3. There is a relationship between student background and biographical characteristics and course performance outcomes as measured by course grade and retention after controlling for test score.	Independent: *Dispositional* High School GPA, High School Education, Years Out of School, Grade in Last English Class Grade in Last Math Class	Assessment Questionnaire merged with Student Transcript and Course Enrollment Database	Multiple Regression for the Course Grade Outcomes Model
	Years of High School English Years of High School Math Highest Level of Math Class How Long Ago was Last Math Class Importance of College to Student Educational Goal, Certainty of Major Veterans Status High School Quality Independent: *Situational* Receive Financial Aid Income Level Units Planned for Next Term Number of Dependents Hours Worked per Week Importance of College to Friends or Family of Subject Marital Status *Dependent*: Course Grade and Retention	Instructor Grade Rosters, and Drop / Withdraw Records	Cross-tabulation and Logistic Regression for the Course Retention Model

The Assessment Questionnaire and Institutional Records

Data from the questionnaire and institutional records were the sources of the biographical variables used in the course performance outcomes model developed in part III of the investigation. The majority of the demographic, situational, and dispositional variables described above were gathered using a questionnaire administered at the time of placement testing. Institutional records were used to supplement the information provided by the assessment questionnaire. Institutional data were gathered from the student's application for admission and financial aid records. These data provided information on the situational characteristics of study participants such as financial aid eligibility, marital status, and the number of dependents.

Demographic Data

The demographic data used in this investigation include several variables found related to performance in college. These data include the sex, race or ethnic background, primary language, income, age, learning disability status, and enrollment status of the subjects (i.e., first time college student, returning student, etc.). Demographic variables are used in the model to control for error introduced by the use of retrospective data in non-experimental conditions as is the case in this study. The demographic variables are shown in Table 5.1 in chapter 5.

Situational Variables

To measure the situational characteristics of students, both proximal and distal data are gathered and employed in this investigation. Proximal data about the situational characteristics of the participants included the time of day the student attended (day or evening). This might be indicative of family or employment responsibilities. Another

situational indicator used was \the number of dependent persons in the student's household. Respondents also provide data on the number of hours per week they are employed while enrolled in college. The self-reported income of the student was also be used as a proximal situational variable that may affect performance or retention in class.

Other situational variables used in part two of this study include the marital status of the student, years since they attended school, and hours spent working off campus each week. These variables were used to measure the situational demands of a student who may have child care needs, job responsibilities, or transportation difficulties that may affect their performance in school.

As suggested in the review chapter, prior research has suggested that the support network of friends or family may be an important variable in whether a student persists in school. To gather this information, participants are asked to indicate how important it is to the people closest to the respondent that he or she attends college. This may be indicative of a situational characteristic pertaining to a support network for the student that affects their motivation and persistence in school (Cross, 1981; Tinto, 1975).

Dispositional Variables

Dispositional variables focus on the motivation, activities, traits, and experiences of the study participants in promoting their own success in school. Dispositional variables include the overall performance and course enrollment patters of the participant while in high school. Dispositional variables are measured by high school graduation, the type of diploma or certificate earned (GED, high school equivalency, foreign diploma), the number of years of high school English and mathematics courses and grades received, and high school grade point average.

Another dispositional characteristic measured in this investigation included the perceived importance of college attendance by the participant. This is intended to gather information on the personal disposition or motivation of students to achieve or excel while in college and the importance they give this activity.

The educational goal of the student was also used as an indicator of a dispositional characteristic. Participant responses to this question provides additional insight into their motivation and goals for college attendance Participants are asked if attending college is to gain personal enrichment, complete a high school diploma, earn a vocational certificate, or eventually attain an associate degree or transfer to a four year college or university.

For part III of the investigation, three databases were used. One database was the college student information system for the three colleges included in this study. This database contains course enrollment history, financial aid information, and a transcript for each student showing all grades received and retention information. The second database used is the assessment questionnaire given at the time of testing. Data from this questionnaire were subsequently merged by the investigator with student test scores and institutional records of the student including transcript information. This merging was completed by electronically aggregating the student enrollment, test score, and course performance data using the unique student identification code as the matching variable. Students were then tracked longitudinally into the course and their grades and retention were gathered from the electronic transcript information. Grades and retention were used as the dependent variables for the course outcomes prediction models developed in part III

Before beginning their placement tests, students were asked to complete the 26 item questionnaire (see Appendix 2 for a copy of the questionnaire) Items taken from the assessment questionnaire are indicated in the table 4.8 and differentiated from data gathered from institutional records such as the student application for admission to the college. Table

4.6 also identifies the variables as dispositional or situational using the scheme provided by Cross. The source of the variables is also identified in table 4.8.

The inclusion of the survey data combined with the student assessment and transcript data in the merged database enabled evidence to be gathered and analyzed to test each null and alternate hypothesis for part III of this study. The use of several independent variables required the use of a multivariate method to analyze the contribution of each block or group of theoretically derived variables to the explanation of variance in course performance.

For part III of this study, two regression models were used depending on the level of measurement of the dependent variable. To explain variance in course grade, multiple linear regression was used to develop the explanatory model. Explaining variance in course retention (a dichotomous variable) was done using a logistic regression model. This approach is recommended in the literature as more appropriate for explaining variance in cases where the dependent variable is binary (i.e., the variable can assume only one of two values) (Norusis, 1992). When the dependent variable can have only two values, the assumptions for hypothesis testing using linear regression analysis may be violated (Hosmer and Lemeshow, 1989).

Logic for the Sequencing of Entering Data in the Model

The course grade prediction model was constructed in a hierarchical approach. Test scores were first entered into the model. Test scores were entered first because in practice, test scores are intended to predict course performance or serve as a proxy for entering ability. The subsequent ordering of the entry of variables into the model was intended to reflect the reality of placement practices in a majority of the state's two year colleges. This reflects both college matriculation policy and institutional practices.

The second block of variables entered into the equation were demographic variables such as race, age, and sex. Demographic variables were entered in this sequence because it is believed that certain demographic characteristics may mediate the predictive power of test scores in predicting course performance. Viewed in this way, the demographic variables may be intervening variables between the placement test score and final course grade. This was confirmed by inspection of the crosstabulation tables. Crosstabulation suggested differential performance on the placement tests by certain demographic categories of students. Crosstabulation also suggested an association between group membership (e.g., older students, females) and the course performance outcomes of final grade and retention. Therefore there was reason to suspect that certain demographic characteristics of the students may help to explain variance in course performance beyond the information provided by testing. Categorical variables such as sex and ethnic grouping were transformed using dummy coding for use in the linear regression and logistic regression models.

The third group of variables entered into the equation were dispositional data. Dispositional variables) may be indicative of more enduring traits about the student. The fourth block of variables entered into the model were situational variables. Situational variables were entered last, because they tend to reflect the more immediate or current circumstances of the student.

The logic for the sequencing of the demographic variables preceding the dispositional and situational variables in the model was that the demographic variables were thought to mediate the situational and dispositional variables. For example, Astin (1991) found that certain demographic variables mediated or affected certain affective or cognitive variables similar to those used in this investigation. Dispositional variables were thought to be more indicative of certain longer-term traits about the student. Dispositional variables

may be viewed as temporally preceding the situational circumstances a student may find themselves in at a given point in their school career. Situational variables may be more of a temporary nature, and can change depending on the vicissitudes of the student's life.

In some instances, the dispositional traits of the student may be related to or affect the situation of the student. For example, a student unsure of their abilities or academically unable to attend college may delay entry into college and choose to work or start a family or both. This may affect their situational circumstances later when they choose to attend school as adults. To reflect the longer-term, more enduring nature of student traits and educational background, the dispositional variables were entered as the second block of variables with the situational characteristics of the student entered as the third and final block of variables.

Missing Data

The use of questionnaire data combined with institutional records and assessment test scores involved the aggregation of several databases. Listwise deletion of missing variables from the two regression models of course grade and retention resulted in some missing cases for part two of the investigation. When listwise deletion is used in a regression model, each case must have complete data for all variables requested or that case is deleted from the model. However the use of listwise deletion is recommended to minimize error in the regression model, and to more fully consider the contribution of each theoretically derived group of variables on the outcomes of interest (Norusis, 1994). In part two of this study , listwise deletion resulted in the elimination of approximately ten to fifteen percent of cases depending on the curriculum level studied. Although there are methods identified in the literature that are used to mitigate the impact of missing data such as substituting the mean for the missing case, or using imputation or weighting the data to

develop predicted observations, none of these methods were used in this study. Measures of central tendency indicated that the distributions of test scores, grades, and student biographical characteristics were approximately normal. There existed a sufficient number of cases such that the reliability of the findings were not significantly threatened. Analysis of missing cases did not reveal a pattern of non-responses from any particular grouping of students or at any particular level in the curriculum. Thus it was reasoned that the sample would not be significantly biased by the deletion of missing values. Therefore it was decided to exclude cases with missing data from the analysis.

Part III: Summary

The research question guiding part III of the study is intended to explain variance in course performance outcomes using demographic, dispositional, and situational variables as independent variables, and final grade and retention as the dependent variables. The extent of data contained in the integrated database enabled the analysis and synthesis of evidence to develop a composite model that to explain variance in the dependent course performance outcomes of final grade and retention. In addition to placement test scores, the inclusion of student biographical variables were used to improve upon the power of the placement test to explain course grades and retention.

Part IV

Faculty Grading Practices as a Source of Variation in College Course Outcomes

The fourth part of the investigation attempted to control for standardized test scores and student biographical characteristics while noting the contribution of instructor grading practices to explaining or predicting variance in course performance outcomes. The model

developed from part IV provided empirical evidence for the combined effect of student test scores, student demographic, dispositional, situational characteristics, and grading variation on the two outcome variables of course grade and retention.

Instructor identification codes were used in the model to detect the extent of instructor-based error associated with the dependent course performance variables. As in part III, the dichotomous dependent variable of retention required the use of a logistic regression procedure to test the instructor effects on retention outcomes.

The research questions guiding part IV were:

> How do teacher characteristics affect the prediction of performance outcomes? Does the addition of instructor information add significantly to the regression R-square (grade prediction) or a significant model chi-square decrease in the -2 log-likelihood (retention prediction) after all other student information has already been entered? Does the employment status (full or part-time employment) of the instructor have a relationship to final grade or retention?

Following the electronic matching of the survey, assessment, and transcript databases conducted for parts II and III of this investigation, the instructors for each course were identified from enrollment records and a review of electronic course schedule information. These instructors were coded using dummy coding and added to the database, each as a separate variable. The questions, coding, and type of question for part IV of this investigation are presented in table 4.9. Table 4.9 also presents a summary of the alternate hypotheses, the independent and dependent variables, source, and analytical approach for part IV of this investigation.

Table 4.9: Alternate Hypothesis, Variables, Source, and Analytical Approach For Part IV of Investigation

Alternate Hypothesis	Variables	Source	Analysis
4. There is a relationship between instructor characteristics and final grades and course retention after controlling for student characteristics and test scores. 4a. Teacher characteristics will account for more variance in the course outcomes of final grade or retention than the simple correlation of test scores and grades or retention. 4b. There is a relationship between instructor status (full- or part-time employment) and student final grade and course retention.	Independent: *Demographic* Variables from Part III Independent *Dispositional* Variables from Part III Independent *Situational* Variables from Part III Independent *Instructor Grading Variation* Instructor Identification Code Instructor Employment Status: 1="Full-time instructor" 2="Part-time instructor (adjunct)" Dependent- Course Grade and Retention	Student Test Score and Demographic Databases Assessment Questionnaire merged with Student Transcript and Course Enrollment Database. Instructor Grade Rosters, and Drop / Withdraw Records	Multiple Linear Regression Logistic Regression

Instructor Codes as Independent Categorical Variables

In both the linear and logistic models used in this investigation, the values for the independent variables must be coded at the appropriate level of measurement to be meaningful. To use the instructor codes in the regression models, instructors were dummy-coded. Using dummy coding, the coefficients represent the effect of each instructor compared to a reference category (Norusis, 1991). The coefficient for the reference

category is zero. Other nominal variables used in the course outcome models such as race, sex, and educational goal were also coded in this way. A more complete discussion of dummy coding can be found in Pedhazur (1982), and Hosmer and Lemeshow (1989).

Analysis

Two models were constructed in part IV of the investigation. The model explaining variance in final grade used the same test score, demographic, dispositional, and situational variables used in part III. However, at the final stage of the regression model, the block of instructor dummy codes were force-entered into the regression. The sequence of entry for each block or category of independent variables into the regression models for English and mathematics is shown in table 4.10 below.

In the case of the model showing the relation of student test scores, biographical data, and faculty grading variation to retention, the change in the significance of the model chi-square decrease in the -2 log-likelihood produced by the logistic regression procedure was monitored at each step of model building. As each block of variables (test score, demographic, dispositional, situational, and instructors) was entered hierarchically into the model, the improvement in significance of the decrease in the -2 log-likelihood of the model provided information as to how well the retention model fit the data, and shows the reduction in error gained by the entry of each block or group of variables included in the model.

To note the relation of faculty employment status to retention, two retention models were produced. One model used only full-time faculty while the other model used part-time. The sequence of variable entry was the same for both models.

Table 4.10: Hierarchical Sequencing of Variables Used in Course Performance Outcomes Model

Category of Independent Variable	Data	
Block One	English Courses	Mathematics Courses
Placement Test Scores	APS Reading and Writing Test Scores	MDTP Algebra Readiness and Pre-Algebra Test Scores
Block Two Demographic Variables	Sex, Race, Age, Primary Language, Learning Disability Status	Sex, Race, Age, Primary Language, Learning Disability Status
Block Three Dispositional Variables	High School GPA, High School Education, Years Out of School, Grade in Last English Class Grade in Last Math Class Highest Degree Earned Years of High School English Years of High School Math Highest Level of Math Class How Long Ago was Last Math Class Importance of College to Student Educational Goal, Units Planned for this Term Certainty of Major	High School GPA, High School Education, Years Out of School, Grade in Last English Class Grade in Last Math Class Highest Degree Earned Years of High School English Years of High School Math Highest Level of Math Class How Long Ago was Last Math Class Importance of College to Student Educational Goal, Units Planned for this Term Certainty of Major
Block Four Situational Variables	Receive Financial Aid Fee exemption status (denotes low income household) Income Level Number of Dependents Hours worked per week Importance of College to Friends or Family of Subject Number of Years Out of School	Receive Financial Aid Fee exemption status (denotes low income household) Income Level Number of Dependents Hours worked per week Importance of College to Friends or Family of Subject Number of Years Out of School
Block Five Instructor Codes	Instructor entered as Dummy Variable Employment Status (Full or Part-Time)	Instructor entered as Dummy Variable Employment Status (Full or Part-Time)

At each step of the regression equation the change in R-squared was observed to determine the relative contribution of each block of variables to the final model. At the final step, the instructor codes were entered into the model to determine the extent of grading variance.

For the model predicting retention, the same approach was taken. To determine the explanatory power of each block of independent variables in retention, the improvement in the model chi-square of the -2 log-likelihood at each step in the development of the model was noted. The improvement is a statistic that tests the null hypothesis that the coefficients for the variables added at the last step are zero. The improvement chi-square test is comparable to the F-change test in multiple linear regression and is produced by SPSS, the statistical package used for this analysis (Norusis, 1994).

Summary

This investigation was constructed upon a sequential testing of hypotheses based on considerations derived from Aptitude-Treatment Interaction theory and Point-to-Point theory. Part I of the analysis provides descriptive data about the participant sample and provides measures of central tendency for the independent and dependent continuous variables used in this investigation. Part II examines the relationship between standardized placement test scores and the dependent course outcome variables of final grade and retention. Part II also compares the outcomes of students who were eligible to enroll in a course to those deemed ineligible for enrollment. Eligibility was determined either through completion of a prerequisite course, or on a score-based inference about their likelihood to succeed in the particular course. Comparison is made using two-by-two contingency tables

to compare the dual outcome of success or non-success, with the dual independent variable of eligibility status (either the student is eligible or not eligible for a course).

Part III of the study uses student biographical gathered from a questionnaire and institutional records to identify traits or characteristics about the student. Responses from the questionnaire were used to measure the demographic, dispositional, situational characteristics of the participants. The primary focus of part III was the relation of test scores, and the situational and dispositional characteristics of the student to the course outcome variables.

Part IV uses student placement test scores from part II, student biographical data from part III, and adds instructor codes to the explanatory model to detect the extent of faculty influence on the final grade or retention behavior of the student. The grouped variables were entered as separate blocks into the course performance outcomes model to partition variance in the dependent variables of final grade and retention.

The ability to control for test score and biographical variables enabled this investigation to provide evidence to either reject or accept Alternate Hypothesis 4 that instructor grading variation contributes a significant amount of error in predicting course outcomes. This will help to better examine the unique contribution of the instructor to the variance in final grade or retention in the course.

CHAPTER FIVE

FINDINGS

The findings of this investigation are presented in this chapter and follow the analytical format and sequence presented in the methodology chapter. Findings from the analysis phase of this investigation are organized by hypothesis with accompanying evidence to either accept or reject the null or alternate hypotheses. This chapter is divided into four parts. Part I provides a demographic profile of the sample of participants used in this investigation. Also included are the responses to the questionnaire used to gather the demographic, dispositional, and situational data from study participants. Following Part I the next three parts correspond to the subsequent three phases or parts of this investigation as described in the methodology chapter.

Part II focuses on the validity of placement testing in predicting or explaining variance in student course performance outcomes. In this section, descriptive data include measures of central tendency for the dependent variables of course grade and retention and for the independent variables of scores on the English and mathematics placement tests. Descriptive data provided in the appendix I include the performance of sub-groups of the participant sample on the standardized placement tests in reading, writing, or mathematics. Data include measures of central tendency including the mean, median, and standard deviation for each group in the sample.

Test score means and standard deviations are presented for the three curricular levels used in this investigation for English and mathematics. Means and standard deviations are presented for students enrolling in college level English courses (transfer level courses), one level below college (Associate degree applicable but non-transferable to a university), and two levels below college (developmental, non-credit courses.) Similarly

for mathematics, measures of central tendency of placement test scores are reported for students enrolling in transfer level mathematics and below transfer level mathematics courses.

Part II also provides the evidence to either accept or reject null Hypothesis 1. Evidence is based on the strength of the predictive validity coefficients between placement test scores, and the course performance outcomes of final grade (GPA1 and GPA2), and retention. Predictive validity coefficients are obtained from the Pearson's Product-Moment correlations between test score and the course outcomes. The magnitude of the observed coefficient is used to determine the statistical significance of the relationship between placement test score and the course outcomes for rejection of the null Hypothesis 1. For the dependent variable of retention, evidence to reject the null Hypothesis 1 is based on the chi-square improvement in the -2 log-likelihood provided by logistic regression.

Pearson correlations between the placement tests scores and final grade are also computed for the three curricular levels of English and mathematics. If students are admitted to a course only if they score above a specific cutting score, there is a restriction of range in the independent variable of placement test score. Although placement during this time was advisory, many students may have chosen to follow enrollment advice and enrolled in courses at a level that corresponded with their placement test score. The potential problem of range attenuation due to the pre-sorting of individuals under an existing placement system was corrected for restriction of range in the coefficient. The corrected r is frequently applied when correlating assessment scores above a specified cut level with course performance measures. In essence, the corrected r is an estimate of what the Pearson r would have been if there was no restriction on the range of data under analysis (State Chancellor's Office, 1991). In this investigation, both the uncorrected and corrected predictive validity coefficients are reported. This analysis presents evidence to

either accept or reject the null Hypothesis 1 as it pertains to the three levels of English or mathematics.

In the case of Alternate Hypothesis 1a positing the correlation coefficient between test score and final grade will fall below the .35 state mandated level, evidence to either accept or reject this hypothesis was based on the correlation coefficient after correcting for restriction of range for the three levels. For the entire group of English and mathematics participants, evidence to accept Alternate Hypothesis 1a was based on the uncorrected coefficient because this represented the performance of the entire group without respect to curriculum level. For both dependent variables of course grade and retention, statistical evidence to either accept or reject the null hypothesis of no relationship between placement test scores and the two dependent variables was based on a probability of at least .05 significance.

Alternate Hypothesis 1b posited that successive scores on a standardized placement test would show a greater point-to-point correspondence stronger than the correspondence between test score and final grade. Evidence to accept Hypothesis 1b was based on the correlation coefficient obtained between the two test scores in comparison with the coefficient between test score and final grade. As described in the methodology chapter, two separate administrations of a standardized placement test were given to a cohort of students after a semester of instruction. Comparing the correlation coefficient for 89 pairs of Reading test scores and 60 pairs of Writing test scores provided strong evidence of a point-to-point correspondence between the two scores and thus to the acceptance of Alternate Hypothesis 1b.

For Alternate Hypothesis 2, (comparison of course outcomes for eligible versus ineligible students), a series of bi-variate contingency tables were used to measure the degree of association between a student's course eligibility status and course outcome. To

determine statistical significance of the effect of the eligibility, the phi statistic was computed. If the probability of the observed phi statistic was less than .05, then the null hypothesis of no difference between the two groups (eligible or not eligible for the course based on attainment of a prerequisite) was rejected.

Alternate Hypothesis 2a stated that the actual difference in course outcomes between the eligible and non-eligible groups of students would be small to modest. Because of a large sample size due to the temporary suspension of mandatory prerequisite enforcement, and the sensitivity of the phi statistic to large sample sizes, statistical significance may have been achieved even though the actual difference in success rates is minimal. To test Alternate Hypothesis 2a, the success rate of the ineligible students as a proportion of the success rate of the eligible students was computed. If ineligible students were found to be 70-80 percent as successful as eligible students, this was sufficient evidence to accept Alternate Hypothesis 2a.

The analyses presented in Part III focus on how predictive validity could be improved by including student biographical information into a model to explain variance in college course performance outcomes. Biographical variables were included in the model using a multivariate analysis that included standardized placement test scores and student demographic, situational, and dispositional characteristics. As each category of variables was entered into the model, the resulting change in R-square (for the linear regression model), or the significance of the improvement of the -2 log-likelihood (for the logistic regression model), was observed to determine the categories of variables that significantly improved the explanatory power of the models.

Part IV of this chapter presents evidence for Alternate Hypothesis 4, 4a, and 4b. Part IV reports the proportion of variance in course performance outcomes explained by teacher differences and grading variation. This analysis quantifies the proportion of

variance explained by the addition of instructor information to the regression R-square (grade prediction) or the model chi-square improvement (retention prediction) after all test scores and other student information has been entered into the model. Separate regressions of grade and retention on student biographical characteristics and instructor codes were also conducted for part-time and full-time instructors. This is intended to address Alternate Hypothesis 4b that focuses on whether the employment status (full or part-time employment) of the instructor has a relationship to final grade or retention.

Part I

Descriptive Data About the Sample

Table 5.1 below presents the demographic characteristics of the sample used in this investigation. The demographic distribution of the participant sample approximately reflects the composition of the total freshman cohort in the three colleges used in this investigation. The data presented below include only students who both tested and enrolled in either an English or mathematics course in the 1994 fall semester.

The participant sample was roughly split between males and females, with a somewhat higher proportion of females in the English courses, with approximately equal representation in the mathematics courses. The racial and ethnic background of participants was approximately the same in both the English and mathematics samples. White, Asian, and Latino/Hispanic students constituted about two-thirds of the participant sample. African-American, Pacific Islander, Filipino, and other non-white students made up the remaining one-third of the sample. Approximately three-quarters of the participant sample spoke English as their primary language. Approximately two-thirds of the participant sample was under the age of 25, while about 15% of the sample were above 31 years of age.

Table 5.1: Demographic Characteristics of Participant Sample for English and Mathematics Course Analysis

Subject	English		Mathematics	
Demographic Variable	Number	Percent	Number	Percent
Sex				
Male	2315	45.0	1987	48.4
Female	2096	49.5	1948	47.5
Unknown	261	4.6	169	4.1
Racial Background				
American Indian	50	.9	41	1.0
Asian	1255	22.3	908	22.1
Pacific Islander	47	.8	28	.7
African American/Black	686	12.2	424	10.3
White	1518	26.9	1104	26.9
Latino/Hispanic	1003	17.8	618	15.1
Filipino	414	7.3	301	7.3
Other Non White	276	4.9	193	4.7
Unknown or Other	388	6.9	487	11.0
Primary Language				
English	4084	72.4	3089	75.3
Other Language	1467	26.0	950	23.1
Unknown	86	1.5	65	1.5
Age				
Below 18	486	8.6	394	9.3
18-21	2927	51.9	2229	52.5
22-25	858	15.2	631	14.9
26-30	543	9.6	397	9.4
31-35	376	6.7	271	6.4
36-40	239	4.2	169	4.0
40 Plus	208	3.7	153	3.6
Verified Learning Disability				
Yes	174	3.1	97	2.4
No	4829	85.7	3383	82.4
Unknown	634	11.2	624	15.2

* N's for sub-groupings may not total to 4,239 for English or 3,719 for mathematics due to missing data from the questionnaire

Dependent Variables

The dependent variables used in this part of the investigation were GPA1, GPA2, retention, and course success. For Hypothesis 1b, reading and writing placement test scores are also included as dependent variables in the correlation of scores achieved by study participants on two separate administrations of the placement test. The proportion

retained is provided in the table for the courses included in the investigation. Measures of central tendency for the dependent continuous variables are presented in table 5.2.

Table 5.2: Descriptive Data for Dependent Variables of GPA and Retention

Course	GPA1	Standard Deviation	GPA2	Standard Deviation	Proportion Retained	N
English						
All Levels	2.34	1.16	1.84	1.40	78.5	3293
College Level	2.57	1.16	1.98	1.48	77.3	846
One Level Below	2.27	1.12	1.77	1.36	78.2	1750
Two Levels Below	2.24	1.22	1.80	1.40	80.7	697
Mathematics						
All Levels	2.70	1.17	2.14	1.51	79.2	2917
College Level	2.82	1.13	2.34	1.48	82.7	937
One Level Below	2.74	1.17	2.20	1.51	80.5	874
Two Levels Below	2.55	1.19	1.92	1.51	75.2	1106

The average for GPA1 for all English courses was 2.34 with a standard deviation of approximately 1.2. The average GPA1 for all mathematics courses was 2.7 with a standard deviation of approximately 1.2. There is a tendency for both measures of grade point average to decline as curricular level drops below college level. Retention showed a different pattern. For all English and mathematics courses, retention was approximately the same at 78%. However, when analyzed by level, retention showed a varying pattern by subject area. With respect to English courses, retention appeared to increase as curriculum levels dropped (77% to 81%), while retention in mathematics courses improved as curriculum levels increased (83% to 75%).

Independent Variables

Placement Test Score Data

APS placement test scores were treated in this study as independent variables in explaining variance in the dependent course outcome variables of final grade and retention. Table 5.3 presents measures of central tendency for the APS Reading and Writing tests. Average test scores and standard deviations are also given by English course level (College Level, One Level Below College, and Two Levels Below College) in the curriculum. For all English courses, the average score on the APS Reading test was approximately 18, with a standard deviation of 6.6 while the mean APS Writing test score was approximately 22 with a standard deviation of 6.4. As might be expected, scores on the placement tests declined in lower levels of the curriculum. As shown in table 5.3, most students tended to follow the enrollment advice with students scoring lower on the placement tests enrolling in pre-collegiate English courses.

Table 5.3: Means and Standard Deviations of APS Reading and Writing Tests

Course	Mean APS Reading Score	Std. Dev.	Mean APS Writing Score	Std. Dev.	Cases
All English Courses	17.7	6.6	21.5	6.4	3925
Transfer Level	24.7	4.9	27.6	5.1	1033
One Level Below	16.8	4.7	20.1	4.8	2113
Two Levels Below	11.8	4.8	15.8	4.9	779

Means and standard deviations for the MDTP (mathematics placement test) are presented for the three curricular levels in table 5.4. The pattern of decreasing scores as curricular levels drop below college level is also evident. As in English, most students appeared to have followed placement advice despite the voluntary, non-restrictive enrollment policy in effect in the participating colleges.

Table 5.4: Means and Standard Deviations of MDTP Algebra Readiness Test

Course	Mean MDTP Algebra Score	Standard Deviation.	Cases
All Mathematics Courses	17.7	7.3	3719
College Level	24.7	4.9	1033
One Level Below College	16.6	5.4	1371
Two Levels Below College	16.9	4.0	1141

Additional independent variables used in this investigation included the demographic, dispositional, and situational characteristics of the study participants. However, these variables are the subject of Hypothesis 3 and are discussed separately in part III of this chapter.

Part II

The Relationship of Standardized Placement Test Scores to Course Performance Outcomes

The relation of placement test score to final grade was determined by calculating the predictive validity coefficient between test score and course grade. To test whether scores on standardized placement tests are related to course retention, a logistic regression was conducted.

Evidence for Hypothesis 1 and 1a is presented in table 5.5 below. The statistical significance was based on the size of the correlation coefficient. For retention, the model chi-square improvement in the -2 log-likelihood was used to test the null hypothesis that the coefficients for the terms in the model except the constant are 0. In the table below, the model chi-square statistic is included with the degrees of freedom for the model explaining retention based on test score.

Table 5.5: Predictive Validity Coefficients for English and Mathematics Placement Tests

Course Level Outcome	Placement Test			N	N
	Reading	Writing	Mathematics	English	Math
All Curriculum Levels					
GPA1	.20**	.26**	.17**	2393	1516
GPA2	.16**	.23**	.16**	3071	1979
Retention Chi-square Improvement in -2 log-likelihood (df)	.093 (1)	12.6** (1)	12.1** (1)	3075	1980
College Level					
GPA1 (corrected)	.22** (.30)	.29** (.36)	.10** (.11)	799	775
GPA2 (corrected)	.19** (.24)	.26** (.33)	.10** (.11)	615	937
Retention Chi-square Improvement in -2 log-likelihood (df)	4.3* (1)	10.1** (1)	1.4 (1)	799	281
One Level Below					
GPA1 (corrected)	.19** (.26)	.25** (.33)	.14** (.15)	1294	567
GPA2 (corrected)	.12** (.16)	.23** (.30)	.15** (.16)	1660	719
Retention Chi-square Improvement in -2 log-likelihood (df)	.02 (1)	17.1** (1)	5.5* (1)	1661	719

Course Level Outcome	Placement Test Reading	Writing	Mathematics	N English	N Math
Two Levels Below	.16**	.23**	.19**	484	731
	(.21)	(.30)	(.26)		
GPA1 (corrected)	.08**	.21**	.15**	612	979
	(.11)	(.28)	(.20)		
GPA2					
	1.4	3.5	8.8**	615	878
	(1)	(1)	(1)		
Retention					
Chi-square					
Improvement in -2 log-likelihood					
(df)					

* p<.05 **p<.01

For all levels of English and mathematics, test scores were found to have a statistically significant relationship ($p < .05$) with the dependent variables of course grade. However with respect to retention, test scores did not always show a statistically significant relationship. More often than not, the relationship between scores on the Reading placement test and retention were not statistically significant. For all levels of English and mathematics the model chi-square improvement in the -2 log-likelihood was not significant for the Reading test, while statistical significance was achieved for the Writing and Mathematics placement tests.

When analyzed by curriculum level, the Reading placement test was significantly related to retention only in College level English courses. However, in College level mathematics courses, mathematics placement test scores were not significantly related to retention. For courses one level below college level, Reading, Writing, and Mathematics placement test scores were significantly related to final grade. With the exception of the Reading placement test, placement test scores were also found to be significantly related to retention at this level.

The relation between test scores and grades in English and mathematics courses two levels below college level was significant. However, retention did not demonstrate a consistent relationship with placement test scores at two levels below college. Only in the case of the mathematics was test score significantly related to retention.

In general, the scores on the Reading placement test showed a weaker relationship with final grades and retention than did scores on the Writing placement test. With the exception of retention in courses one level below college, the Mathematics placement test was significantly related to retention and final grade.

In total, a preponderance of the evidence showed a statistically significant relationship between placement test scores and the dependent variables, leading to the rejection of null Hypothesis 1 and the acceptance of the alternate Hypothesis 1. There was sufficient evidence that a statistically significant relationship existed between placement test scores and final grade and retention.

The alternate Hypothesis 1a posited that although test scores may show a statistically significant relationship to final grade, the coefficient will fall below the .35 level specified by California educational code as the legal minimum for test validation. Inspection of table 5.5 shows that the coefficients did not achieve the .35 level for all levels of English and mathematics. When examined by curricular level, the .35 coefficient was achieved only in the case of college level English courses after correcting for restriction of range. The low to modest correlation coefficients found in table 5.5 led to the acceptance of Alternate Hypothesis 1a. and the rejection of the Null Hypothesis 1a.

Alternate Hypothesis 1b tested if scores on successive administrations of a placement test showed a greater point-to-point relationship (i.e., a higher correlation coefficient) than the coefficient between test score and final grade. Comparison of pairs of scores for 89 students taking the reading placement test and for 78 students taking the

writing placement test suggests that the pairs of placement test scores are highly related as shown in table 5.6 below.

Table 5.6: Correlation of Reading Scores in Fall 1997 with Reading Scores in Spring 1998 and Writing Scores in Fall 1997 with Writing Scores in Spring 1998

Test	Reading Spring, 1998	Writing Spring, 1998	Prob. (2-tailed)	N
Reading Fall, 1997	.65*		.000	89
Writing Fall, 1997		.78*	.000	60

*p < .001

The correlation coefficient between the two scores was superior to the coefficient observed between test score and final course grade. The correlation coefficient between successive administrations of both the reading and writing placement test far exceeded the value observed between placement tests and final grades as shown in table 5.5, even after correcting for restriction of range in the placement test scores. The findings of a higher correlation coefficient between test scores compared to the coefficient for test scores and final grade led to the rejection of Null Hypothesis 1b and to the acceptance of Alternate Hypothesis 1b.

Further analysis also revealed the relatively high degree of statistical reliability between common measures. For example, after a semester of instruction, scores on the placement test were not found to differ significantly. Comparing the mean scores using a matched pairs t-test shows no statistical difference between the two pairs of scores. This was confirmed by a t- value of 1.09 for the reading test and -.42 for the writing test as shown in table 5.7.

Table 5.7: Average Test Scores of a Cohort of Students on Two Successive Administrations of Reading and Writing Placement Tests

Test	Average Score Fall, 1997 (Std. Dev.)	Average Score Spring, 1998 (Std. Dev.)	t value	Prob. (2-tailed)	N (pairs of scores)
Reading	23.8 (7.8)	23.1 (6.8)	1.09	.28	89
Writing	23.5 (6.6)	23.7 (6.4)	-.42	.68	78

Further analysis was conducted with students who had taken a placement test at the start of the fall semester, completed an English course successfully (i.e., earned a grade of 'C' or higher), and enrolled in a higher level English course. This analysis was conducted to reduce the number of students who may have taken an English course but performed poorly and had dropped a level or were repeating the course. The minimal difference in scores observed in the table 5.7 over one semester may have been due to the inclusion of several students who did not perform well in their English class during the fall semester. Students successfully completing their English class and enrolling at least one level higher were tracked longitudinally to determine if their test scores improved on a second administration of the placement test. The results of this analysis are shown in table 5.8.

Table 5.8: Comparison of Placement Test Scores for Students Advancing One Level in the English Curriculum

Test	Average Score Fall, 1997 (Std. Dev.)	Average Score Spring, 1998 (Std. Dev.)	t value	Prob. (2-tailed)	N
Reading	23.5 (7.3)	22.7 (6.2)	.93	.35	60
Writing	23.0 (5.3)	24.1 (5.4)	-1.95	.06	46

This analysis also showed that even for students who passed their English course in the fall term with a grade of 'C' or better and advanced to a higher level English course, there were no statistically significant difference between pre- and post-test scores.

The Relation of Student Course Eligibility to Course Success

Alternate Hypothesis 2 stated:

There is a relationship between course eligibility status (i.e., a student is either eligible or eligible for enrollment in a particular course) and success in the subsequent selected community college courses.

Course eligibility was the independent variable and course grades were used as the dependent variable in this phase of the study. Success in the course was defined as attaining a grade of 'C' or better. Some courses required a communication prerequisite (such as college level English or an equivalent placement test score), others required a computational prerequisite (such as elementary algebra or an equivalent placement test score), while some required both a communication and a computation prerequisite.

Table 5.9: Comparison of Performance for All Students by Eligibility Status

Subject All Courses	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
English Skill						
Eligible	32	68				434566
Not Eligible	41	59	-.06	***	87%	54770
Math Skill						
Eligible	33	67				101403
Not Eligible	46	54	-.11	***	81%	30640

* p< .05 ** p<.01 *** p<.001

Comparison of the success rates of students based on course eligibility for all courses in the three community colleges requiring either a communication (English skill) or computation (Math skill) prerequisite is shown in table 5.9. Students eligible for the course were more likely to earn a grade of 'C' or higher than were students without the recommended prerequisite. The phi statistic was of sufficient size to reject the null

Hypothesis 2 of no relation between course eligibility and success, and to accept the Alternate Hypothesis 2 of a relationship between course eligibility and course success.

The Alternate Hypothesis 2a posited that the actual difference between the success rates of the eligible and ineligible students would be modest. This was confirmed by computing the relative success rate of the ineligible to the eligible students. For example it was found that for all courses requiring an English skill level prerequisite, 68% of students meeting the prerequisite were successful while 59% of those without the prerequisite who attempted the course were successful. Although these differences were statistically significant, when the success rate of the ineligible is computed as a proportion of those eligible, it was found that the ineligible were 87% as successful as those considered eligible or qualified to attempt the course.

With respect to course eligibility based on a pre-determined mathematics skill, participants without the required mathematics skill for a particular course were 81% as successful as participants with the required prerequisite skill level. When analyzed in this manner, the differences do not appear to be great. This evidence was used to reject Null Hypothesis 2a and accept the Alternate Hypothesis 2a. The practical difference in success rates for eligible compared to ineligible students is small to modest.

Comparison of Performance by Subject Area

Tables 5.10 through 5.16 show similar results for the several subject areas selected for analysis. Table 5.10 compares outcomes for students enrolled in accounting courses requiring a prerequisite.

Table 5.10: Comparison of Performance for Accounting Students by Eligibility Status

Subject Accounting	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
English Skill						
Eligible	48	52				2081
Not Eligible	58	43	-.04	**	83%	160
Math Skill						
Eligible	46	54				1931
Not Eligible	67	33	-.13	***	61%	238

* p< .05 ** p<.01 *** p<.001

The accounting courses required both an English and a mathematics skill level or prerequisite for entry. As might be expected, attainment of the pre-determined mathematics skill level or prerequisite was more strongly related to course success than was attainment of the English prerequisite skill level. The relationship as measured by the phi statistic was sufficient to merit rejection of the null hypothesis of no relation between course eligibility and success with respect to accounting courses. However it can be noted that with respect to the English skill level prerequisite, non-eligible students who under a mandatory placement system would have been denied entry into the course, were 83% as successful as eligible students. Non-eligible students on the basis of a mathematics prerequisite were 61% as successful as those with the prerequisite. The mathematics prerequisite appears to have a stronger empirical relation to actual course success even though approximately one-third of students without this demonstrated skill level were successful.

Table 5.11 shows similar results for chemistry courses. As with the accounting courses described above, chemistry courses also required an English and mathematics prerequisite skill level. The relation between course eligibility and success based on attainment of the English prerequisite was not significant. Interestingly, students without

the required English skill level were slightly more successful than students with the prerequisite skill level. This was confirmed by the small value of the phi statistic and the finding that the non-eligible were 104% as successful as the eligible students. Attainment of the mathematics prerequisite was also found to be statistically unrelated to course success.

Table 5.11: Comparison of Performance for Chemistry Students by Eligibility Status

Subject Chemistry	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
English Skill						
Eligible	38	62				1530
Not Eligible	35	65	.01	NS	104%	298
Math Skill						
Eligible	37	63				1778
Not Eligible	43	57	-.02	NS	75%	65

* p< .05 ** p<.01 *** p<.001

Table 5.12 compares outcomes for students enrolling in economics courses. For Economics students, attainment of the English skill level prerequisite was found to be significantly related to success. This was confirmed by the value of the phi statistic (p <.001). However, just over one-half of students attempting the course without the prerequisite were successful. It was found that students without the prerequisite were approximately 79% as successful as those with the prerequisite.

Table 5.12: Comparison of Performance for Economics Students by Eligibility Status

Subject Economics English Skill	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
Eligible	34	66				3279
Not Eligible	48	52	-.12	***	79%	963

* p< .05 ** p<.01 *** p<.001

The two-way contingency table comparing outcomes for engineering students on the basis of course eligibility is presented below. Among those eligible for the engineering course, 65% earned a grade of 'C' or greater while among those considered ineligible, 53% earned a grade of 'C' or better. Thus course eligibility was significantly associated with course outcomes as shown by the phi statistic ($p<.01$). However, non-eligible students were approximately 82% as successful as the eligible students. As with the other disciplines analyzed, the rationale for excluding students who are approximately 80% as successful as the group of eligible students suggests little practical difference in the actual outcomes.

Table 5.13: Comparison of Performance for Engineering Students by Mathematics Eligibility Status

Subject Engineering	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
Math Skill						
Eligible	35	65				137
Not Eligible	47	53	.12	**	82%	79

* p< .05 ** p<.01 *** p<.001

Table 5.14 compares outcomes for students by curriculum level in English. Three analyses are displayed in table 5.14. Outcomes for students in pre-collegiate English writing and reading courses based on eligibility status show a significant association with attainment of the prerequisite. Approximately 48% of eligible students were successful in pre-collegiate writing courses compared with a 37% success rate of non-eligible students. Similar results were found for pre-collegiate reading courses, although slightly more than one-half of the non-eligible students were successful in the course. Results for college level English courses show a significant relation between eligibility status and success. However, non-eligible students were approximately 75% as successful as eligible students.

Table 5.14: Comparison of Performance for Pre-Collegiate and College Level English Students by Eligibility Status

Pre-Collegiate English Writing	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
Eligible	52	48				5258
Not Eligible	63	37	-.07	***	77%	891
Pre-Collegiate English Reading						
Eligible	37	63				5725
Not Eligible	47	53	-.07	***	84%	1181
College Level English						
Eligible	40	69				10834
Not Eligible	49	51	-.08	***	74%	3987

* p< .05 ** p<.01 *** p<.001

Table 5.15 compares the performance of eligible and non-eligible students in introductory history courses. For entry into this course participants were required to have a college level English skill level as indicated by passing the Reading and Writing placement tests or passing an English course. Among those eligible for the course, 72% earned a grade of 'C' or higher, while among those ineligible, 56% earned a grade of 'C' or higher. The phi statistic indicates that this is a statistically significant difference in performance. However, the relative success of the ineligible is approximately 78% of those eligible. Although there is a statistically significant association between prerequisite attainment and success, the practical differences do not appear to be great, particularly when students denied access had over a 50% chance of success in the course.

Table 5.15: Comparison of Performance for History Students by Eligibility Status

Subject History	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
English Skill						
Eligible	28	72				3310
Not Eligible	44	56	-.12	***	78%	567

* p< .05 ** p<.01 *** p<.001

Table 5.16 compares the outcomes for students based on eligibility status for office information systems courses. These courses are vocational courses designed to teach skills in office technology and related software applications. These courses required both an English and a mathematics prerequisite skill level. With respect to the English skill level prerequisite, there was no reliable relationship noted between attainment of the English prerequisite and course success. This was confirmed by the low value of the phi statistic and the finding that the non-eligible on the basis of English prerequisite attainment were more successful than those who possessed the prerequisite. Attainment of the mathematics prerequisite did demonstrate a statistically significant relationship with success (p<.05). However, 61% of those without the mathematics prerequisite were able to earn a grade of 'C' or higher. Further analysis showed that those without the mathematics prerequisite were 85% as successful as those with the prerequisite.

Table 5.16: Comparison of Performance for Office Information Systems Students by English Eligibility Status

Subject Office Information Systems	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
English Skill						
Eligible	33	67				313
Not Eligible	27	73	.04		108%	66

Subject	Percent Not Successful	Percent Successful	Phi	Sig.	Relative Success of Non-Eligible Compared to Eligible	N
Office Information Systems						
Math Skill						
Eligible	28	72				272
Not Eligible	39	61	-.10	*	85%	117
* p< .05	** p<.01		*** p<.001			

Alternate Hypothesis 2 regarding a relationship between eligibility status and course outcomes was supported by the data. With respect to Alternative Hypothesis 2a, the evidence suggests that ineligible students are approximately 70-80% as successful as their eligible peers for most of the subject areas analyzed. The Alternate Hypothesis 2a of minimal practical difference in outcome between eligible and non-eligible students was supported by the evidence and replicated among students in the seven disciplines analyzed.

Part III

Course Performance Outcomes Model

The third part of the investigation focused on the development of a theoretical model to improve understanding of student performance and retention in community college English and mathematics courses. The null Hypothesis 3 that guided Part III of the study was:

> There is no relationship between student background and biographical characteristics and course performance outcomes as measured by course grade and retention after controlling for test score.

The Relationship of Demographic Characteristics and Course Performance Outcomes

The development of the theoretical model for final grade and retention included an analysis of available demographic data to determine the degree of association between student characteristics and the course outcomes of final grade and retention. The statistical significance of the association between demographic characteristic and course outcomes

was determined by using the chi-square statistic. Demographic, dispositional, and situational variables found to have a statistically reliable association with the dependent outcome variable of success are noted in the row labeled "significance" in table 5.17. The independent demographic data included in table 5.17 are the variables entered into the demographic block of variables in the course performance outcomes model later in this section.

Table 5.17: Crosstabulation of Demographic Characteristics and Course Performance Outcomes

	English		Mathematics			
Demographic Characteristic	Percent Successful (A, B, C, Credit)	Percent Retained	Percent Successful (A, B, C, Credit)	Percent Retained	English (N)	Math (N)
Race/Ethnic						
American Indian	57	63	44	56	(35)	(25)
Asian	65	83	78	90	(552)	(627)
Pacific Islander	65	72	67	71	(29)	(24)
Black	51	73	47	66	(462)	(325)
White	63	77	69	77	(983)	(924)
Latino/Hispanic	55	77	58	74	(638)	(494)
Filipino	65	87	74	87	(305)	(248)
Other	58	78	57	75	(125)	(150)
Significance	**	**	**	**		
Sex						
Male	57	77	66	79	(1480)	(958)
Female	62	79	67	79	(1671)	(942)
Significance	**					
Age						
Below 18	65	80	69	81	(332)	(334)
18-21	58	80	65	80	(1877)	(1650)
22-25	60	75	73	80	(435)	(468)
26-30	64	76	69	76	(260)	(316)
31-35	68	78	70	79	(173)	(195)
36-40	59	75	71	79	(113)	(128)
40 Plus	61	80	77	85	(108)	(115)
Significance			**			
English Primary Language						
Yes	59	76	63	76	(2167)	(1838)
No	61	82	74	85	(950)	(971)
Significance		**	**	**		

	English		Mathematics			
Demographic Characteristic	Percent Successful (A, B, C, Credit)	Percent Retained	Percent Successful (A, B, C, Credit)	Percent Retained	English (N)	Math (N)
Learning Disabled						
Yes	45	73	49	68	(113)	(65)
No	61	78	67	79	(2925)	(2645)
Significance	**		**	*		

* Denotes significance at .05 ** Denotes significance at .01

The race or ethnic background of the participant was significantly related to both dependent variables of final course grade and retention in English and mathematics courses. Females were more likely to earn passing grades in English compared to males ($p<.01$). There were no statistically significant differences found between males and females in English course retention, mathematics course grade and mathematics course retention. The age of the participant was significantly associated with success in mathematics courses. Older students showed somewhat higher success rates than younger students. Age was not found to be statistically related to course outcomes in English courses or retention in mathematics courses. An interesting finding was that for participants indicating a language other than English as their primary language, the retention in English courses, and both grades and retention in mathematics courses were significantly higher than those whose primary language was English. Participants identifying themselves as having a verified learning disability tended not to be as successful as students without a learning disability. The learning disability status of the participant was not found to be significantly related to retention in English courses. To more fully analyze the contribution of demographic variance to course performance outcomes, the demographic variables listed above were included in the regression models for final grade and retention.

The Relation of Situational Variables to Course Performance Outcomes

Table 5.18 presents a summary of the crosstabulation of situational variables with the course performance outcome variables. As with the crosstabulation of the demographic variables above, the significance of the association between dependent and independent variables was determined using the chi-square statistic.

Table 5.18: Crosstabulation of Situational Variables with Course Performance Outcomes

Situational Characteristic	English Percent Successful (A, B, C, Credit)	Percent Retained	Mathematics Percent Successful (A, B, C, Credit)	Percent Retained	English (N)	Math (N)
Time of Attendance						
Day	59	79	64	78	(1838)	(1581)
Evening	62	76	71	79	(475)	(484)
Both Day and Evening	63	78	69	81	(757)	(667)
Significance			*			
Units Planned Next Term						
Less Than 6	63	79	69	79	(377)	(300)
6 to 8	59	78	65	77	(587)	(502)
9 to 11	56	73	64	79	(412)	(374)
12 or More	61	79	68	79	(1547)	(1414)
Significance						
Employment Hours Per Week						
None	59	78	71	82	(759)	(628)
1-10	63	82	68	79	(202)	(194)
11-20	63	83	64	79	(720)	(641)
21-30	58	75	63	78	(627)	(554)
31-40	63	75	68	78	(494)	(433)
More Than 40	60	76	69	78	(151)	(129)
Significance		**				
Number of Dependents						
None	60	90	67	80	(2985)	(2091)
1-2	60	87	65	75	(403)	(509)
3 or More	58	87	68	82	(189)	(277)
Significance		**		*		

Situational Characteristic	English Percent Successful (A, B, C, Credit)	Percent Retained	Mathematics Percent Successful (A, B, C, Credit)	Percent Retained	English (N)	Math (N)
Importance of College to People Closest to You						
Not Very Important	58	72	75	83	(144)	(133)
Somewhat Important	61	78	66	79	(671)	(562)
Very Important	60	78	65	78	(2174)	(1903)
Significance		*	*			
Receive Financial Aid						
Yes	61	81	65	80	(1190)	(1669)
No	63	78	70	80	(2142)	(1494)
Significance						
Marital Status						
Single	61	79	67	80	(3184)	(2444)
Married	67	82	74	80	(599)	(426)
Significance	**		**			

* Denotes significance at .05 ** Denotes significance at .01

Compared with demographic variables, situational variables did not tend to show a statistically significant relationship with the dependent variables. However, many of the direction of the values of the variables did suggest a linear relationship with the dependent variables even though the chi-square statistic was not significant. For example participants who attend courses in the evenings tended to show somewhat higher success rates in both English and mathematics (as measured by course grades) than students attending in the daytime. Retention was approximately the same for day and evening students in English, while in mathematics, evening students showed somewhat higher retention rates. The number of units planned by the participants for the next term did not show a directional relationship with the dependent variables either in grades or retention. Participants who were employed more hours per week had significantly lower retention rates in English

courses than those who worked fewer hours per week. There was a modest improvement in retention rates in mathematics for participants who worked less than 20 hours per week. Employment hours do not appear to have a significant association with grades earned in English or mathematics courses.

It appears that retention was more strongly affected by the situational characteristics of the students than was final grade. For example, retention was lower for English students who worked more hours per week. Students with more family responsibilities as suggested by the number of dependents were more likely to drop out than were students with fewer dependents. An increase in the number of dependents in the household was negatively related to retention in English courses. The importance of important peers or family members on the course performance or retention outcomes of participants was significant for retention in English courses, and final grade in mathematics courses..

Marital status was significantly related to course grades in both English and mathematics, with married students tending to have a higher proportion of successful grades. Retention was also somewhat better for married students in English courses, although the differences were not statistically significant. With respect to grades, the number of dependents did not show a significant association.

The Relationship of Dispositional Characteristics to Course Performance Outcomes

Most of the student dispositional characteristics showed a strong relationship with course grade, retention, or both. Dispositional variables demonstrating the most statistically reliable relationships with the dependent variables included whether the student finished high school and the type of secondary degree earned, the number of years out of school, high school grade point average, the highest level of mathematics completed, the number of years of high school English or mathematics, and grades received in the

participant's last English or mathematics class. The importance of college to the participant showed a reliable relationship to course grade received in English. The educational goal and the certainty of the choice of major of the participants did not appear to be significantly related to the dependent variables.

The quality of the high school attended by the participant was found to be significantly related to retention in college mathematics courses. Students from lower quality high schools (as determined by the local school district) were more likely to drop out of mathematics courses than were students from average or above average high schools.

Table 5.19: Crosstabulation of Dispositional Characteristics with Course Performance Outcomes

Dispositional Characteristic	English Percent Successful (A, B, C, Credit)	Percent Retained	Mathematics Percent Successful (A, B, C, Credit)	Percent Retained	English (N)	Math (N)
High School Education						
Still in High School	61	83	68	83	(501)	(439)
Did Not Complete High School	44	62	50	65	(125)	(105)
High School Diploma	61	79	66	79	(2099)	(1812)
Foreign Diploma	66	82	77	85	(151)	(231)
GED	50	66	61	74	(168)	(131)
High School Proficiency Test	66	75	62	69	(32)	(29)
Certificate of Completion	65	85	59	73	(20)	(22)
Significance	**	**	**	**		
Years Out of School						
Still in School	61	81	68	82	(1063)	(965)
Less Than 1 Year	55	76	61	78	(756)	(601)
1-2 Years	55	71	62	72	(304)	(300)
3-4 Years	61	76	68	78	(250)	(242)
5-10 Years	64	77	72	79	(376)	(354)
More Than 10 Years	65	80	70	78	(306)	(277)
Significance	**	**	**	*		

Dispositional Characteristic	English Percent Successful (A, B, C, Credit)	Percent Retained	Mathematics Percent Successful (A, B, C, Credit)	Percent Retained	English (N)	Math (N)
Years of High School English						
Less Than 1 Year	61	77	81	88	(137)	(169)
One	59	76	68	84	(181)	(152)
Two	57	76	62	71	(270)	(235)
Three	57	76	65	78	(500)	(459)
Four	61	79	64	78	(1924)	(1615)
Significance			**	**		
Grade in Last English Class						
A	72	83	75	86	(440)	(447)
B	64	81	68	79	(1179)	(1022)
C	55	75	59	75	(1012)	(791)
D	41	72	51	71	(159)	(124)
F	40	60	61	67	(15)	(18)
Significance	**	**	**	**		
High School Grade Point Average						
3.5-4.0	78	88	85	91	(303)	(331)
3.0-3.4	67	83	73	83	(639)	(639)
2.5-2.9	59	78	65	78	(874)	(740)
2.0-2.4	56	75	55	72	(769)	(587)
1.5-1.9	49	72	52	67	(309)	(217)
1.0-1.4	33	59	56	72	(54)	(43)
0.0-0.9	25	67	64	64	(12)	(14)
Significance	**	**	**	**		
Highest Level Math Completed						
None	50	46	77	84	(28)	(57)
Basic Math	75	55	60	75	(422)	(321)
Beginning Algebra	70	51	57	70	(659)	(501)
Geometry	78	58	62	78	(488)	(393)
Intermediate Algebra	79	61	65	79	(736)	(677)
Trigonometry	86	71	75	86	(349)	(334)
College Algebra	87	71	81	87	(232)	(268)
Calculus	89	75	81	86	(105)	(110)
Significance	**	**	**	**		

Dispositional Characteristic	English Percent Successful (A, B, C, Credit)	Percent Retained	Mathematics Percent Successful (A, B, C, Credit)	Percent Retained	English (N)	Math (N)
Grade Received in Last Math Class						
A	67	80	79	87	(432)	(415)
B	61	78	70	81	(845)	(773)
C	61	78	59	74	(1153)	(938)
D	47	75	58	76	(429)	(312)
F	67	80	61	79	(51)	(38)
Significance	**	**	**	**		
How Long Ago Last Math Class						
Currently Enrolled	63	85	68	83	(509)	(458)
Less Than 1 Year	58	79	64	79	(862)	(715)
1-2 Years	58	74	61	76	(609)	(525)
3-5 Years	60	77	68	77	(378)	(335)
More Than 5 Years	63	76	70	78	(652)	(595)
Significance		**	**			
Most Important Educational Goal						
Personal Enrichment	38	65	78	85	(60	(54)
High School Diploma	50	80	57	83	(60	(53)
Vocational Training	59	74	62	79	(113	(112)
Associate Degree	57	75	67	78	(530	(421)
Transfer to 4 Year	62	80	66	79	(2023	(1805)
Other	60	76	66	76	(168	(147)
Significance	**	*				
How Definite Choice of Major						
Very	59	78	63	78	(936)	(788)
Fairly	61	78	68	79	(917)	(740)
Unsure	62	79	68	79	(826)	(739)
Significance						
How Important is College to You Personally						
Not Very	35	70	81	94	(20)	(16)
Somewhat	55	75	66	82	(224)	(208)
Very	61	78	66	78	(2737)	(2377)
Significance	*					
High School Quality						
Meets Standards	61	78	68	80	(2688)	(2711)
Below Standards	57	79	64	80	(610)	(465)
Significance			*			

Course Performance Outcomes Model

The contribution of each block of variables to explaining variance in course grade in English and mathematics courses is shown in tables 5.20 and 5.21. After entering test scores, the remaining blocks of variables were entered into the regression. As each block was entered, the change in R-square (proportion of variance explained) was noted. As R-square increased, the significance of the R-square change was determined by the F statistic at each stage of the regression.

In the case of explaining variance in English course grades, the test score, demographic, dispositional, and situational blocks of variables contributed significantly to change in R-square. Of the student biographical variables entered, test score, demographic, and dispositional blocks of variables tended to result in greater proportional reductions in error than situational variables. Dispositional variables tended to explain the most variance in final grade. Dispositional variables also showed a stronger statisitical relationship with course grade than placement test score.

Table 5.20: All English Courses; Explaining Variance in Course Grade

Block	Multiple R	R Squared	F	Significance of Change
Placement Test Scores	.25	.06	37.9	p<.001
Demographic	.36	.13	14.3	p<.001
Dispositional	.45	.20	8.6	p<.001
Situational	.46	.21	7.2	p<.001

When examined for all mathematics courses, placement test scores did not demonstrate a reliable relationship with final grade at the .05 level. However, demographic, dispositional, and situational variables showed a statistically significant relationship with final grade. Of the categories of variables, demographic and dispositional

variables explained a proportionately greater amount of variance in final grade than either placement test score or situational variables.

Table 5.21: All Mathematics Courses; Explaining Variance in Course Grade

Block	Multiple R	R Squared	F	Significance of Change
Placement				
Test Scores	.05	.002	1.78	p> .10
Demographic	.38	.138	9.50	p<.000
Dispositional	.45	.202	4.87	p<.000
Situational	.49	.240	4.75	p<.000

After controlling for test scores, the addition of demographic, dispositional, and situational blocks of variables explain a significant amount of variance in course grade in mathematics courses. Of the variables analyzed, demographic and dispositional variables explained a proportionately greater amount of variance than test scores or situational variables.

Retention

With respect to retention in English courses, test score, demographic, and dispositional variables demonstrated a significant relationship to retention, while the situational variables were not significant.

Table 5.22: All English Courses; Logistic Regression Retention Model

Block	Model Chi-Square	-2 Log - Likelihood Change	Significance Chi-Square Improvement
Placement			
Test Scores	9.8	1663.6	p<.008
Demographic	30.1	1633.2	p<.000
Dispositional	42.2	1590.9	p<.006
Situational	9.1	1530.5	p>.20

With respect to mathematics courses, demographic and dispositional characteristics were significantly related to retention. This is evident by the significant change in the -2 log-likelihood as demographic and dispositional of variables are added to the model. Placement test scores and situational variables did not demonstrate a statistically significant relationship with retention in mathematics courses.

Table 5.23: All Mathematics Courses; Logistic Regression Retention Model

Block	Model Chi-Square	-2 Log-Likelihood	Significance Chi-Square Improvement
Placement Test Score	2.7	971.6	p<.10
Demographic	27.2	944.4	p<.003
Dispositional	46.6	897.8	p<.002
Situational	11.0	886.8	p>.20

The null Hypothesis for part III of this investigation posited no relationship between student biographical variables and the course performance outcomes of final grade or retention after controlling for test score. The data in tables 5.20 through 5.23 however provide sufficient evidence to reject the null hypothesis and accept the Alternate Hypothesis 3. In particular, student dispositional characteristics tended to contribute significantly to the explanatory model as evidenced by the significant change in R-square.

For both English and mathematics courses, dispositional variables tended to have more explanatory power for course grade than placement test scores. With respect to retention, the demographic and dispositional variables contributed to more significant reductions in the -2 log-likelihood of the logistic model than other blocks of variables. These findings suggest the potential usefulness of dispositional and demographic variables in explaining student retention in English and mathematics courses.

Part IV

Instructor Characteristics as a Source of Variance

The fourth part of the investigation controlled for standardized placement test scores and student dispositional and situational characteristics while noting the contribution of instructor effects in explaining variance in the course performance outcomes of final grade and retention. The model developed in part IV provides empirical evidence for the combined effect of student test scores, demographic, dispositional, situational characteristics, and grading variation on the two outcome variables of course grade and retention. In part IV, instructor identification codes were used in the model to detect the extent of instructor-based error associated with explaining or predicting course grade and retention. The hypotheses guiding this part of the investigation are restated below.

> There is no relationship between instructor characteristics and final grades and course retention after controlling for student characteristics and test scores.
>
> Teacher characteristics will account for more variance in the course outcomes of final grade or retention than the simple correlation of test scores and grades or retention.

The next four tables include the contribution of grading variation to the dependent variables for all levels of English and mathematics courses. Table 5.24 notes the change in the R-square after the addition of the instructor block to the model. The addition of the instructor adds significantly to the change in R-square ($p<.001$) even after controlling for all other variables in the model. This is evidence of significant instructor effects on the outcome variable of final grade in English courses.

Table 5.24: All English Courses ; Explaining Variance in Course Grade

Block	Multiple R	R Squared	F	Significance of Change
Placement				
Test Scores	.25	.06	37.9	p<.001
Demographic	.36	.13	14.3	p<.001
Dispositional	.45	.20	8.6	p<.001
Situational	.46	.21	7.2	p<.001
Instructor ID	.60	.36	3.8	p<.001

When examined for all mathematics courses as shown in table 5.25, the addition of the instructor codes added significantly to the change in R-square and contributed significantly to the explanatory power of the course grade model. Instructor codes resulted in a change in the proportion of variance explained of approximately 20%.

Table 5.25: All Mathematics Courses ; Explaining Variance in Course Grade

Block	Multiple R	R Squared	F	Significance of Change
Placement				
Test Scores	.05	.002	1.78	p> .10
Demographic	.38	.138	9.50	p<.000
Dispositional	.45	.202	4.87	p<.000
Situational	.49	.240	4.75	p<.000
Instructor ID	.67	.453	3.24	p<.000

Instructor Effects on Retention

Table 5.26 suggests the significant contribution of the instructor to retention in English courses. The statistically significant change in the -2 log-likelihood after controlling for test scores, demographic, dispositional, and situational variables introduced by instructor identification codes provides evidence of the importance of the instructor in affecting the retention of students.

Table 5.26: All English Courses; Logistic Regression Retention Model

Block	Model Chi-Square	-2 Log - Likelihood Change	Significance Chi-Square Improvement
Placement			
Test Scores	9.8	1663.6	p<.008
Demographic	30.1	1633.2	p<.000
Dispositional	42.2	1590.9	p<.006
Situational	9.1	1530.5	p>.20
Instructor ID	156.8	1425.0	p<.003

In mathematics courses, the addition of instructor variables also contributed significantly to change in the -2 log-likelihood as shown in table 5.27. This is evidence of significant instructor effects on the retention of students in mathematics courses.

Table 5.27: All Mathematics Courses; Logistic Regression Retention Model

Block	Model Chi-Square	-2 Log- Likelihood Change	Significance Chi-Square Improvement
Placement			
Test Score	2.7	971.6	p<.10
Demographic	27.2	944.4	p<.003
Dispositional	46.6	897.8	p<.002
Situational	11.0	886.8	p>.20
Instructor ID	138.9	747.9	p<.005

The evidence provided in tables 5.24 through 5.27 is sufficient to reject null Hypothesis 4, and accept the alternate hypothesis of a significant effect of instructor characteristics on course grade and retention. The addition of the instructor block of variables contributed more explanatory power than placement test scores, thus the Null Hypothesis 4a is rejected and the Alternate Hypothesis 4a is accepted.

Explaining Variance in Course Outcomes by Curriculum Level

To further analyze the contribution of placement test scores, demographic, dispositional, situational, and instructor variables on course outcomes, explanatory models were produced for the three curriculum levels in English and mathematics. Table 5.28

presents separate analyses by English curricular level for the test score, demographic, dispositional, situational, and instructor variables included in the explanatory model for course grade.

Table 5.28: Course Performance Model; Final Course Grade in English by Curriculum Level

Level	Variables	Multiple R	R Squared	F	Significance of Change
College					
	Placement Test Scores	.29	.08	17.77	p<.000
	Demographic	.41	.16	6.10	p<.000
	Dispositional	.57	.32	4.84	p<.000
	Situational	.59	.35	4.36	p<.000
	Instructor ID	.69	.48	3.21	p<.000
One Level Below (pre-college)					
	Placement Test Scores	.28	.08	27.93	p<.000
	Demographic	.37	.14	8.50	p<.000
	Dispositional	.47	.22	5.18	p<.000
	Situational	.49	.24	4.61	p<.000
	Instructor ID	.64	.42	3.77	p<.000
Two Levels Below (basic skills)					
	Placement Test Scores	.22	.05	5.60	p<.005
	Demographic	.47	.22	5.08	p<.000
	Dispositional	.54	.30	2.50	p<.000
	Situational	.59	.35	2.54	p<.000
	Instructor ID	.70	.50	2.70	p<.000

Analyzed by curricular level, all categories of variables entered into the model demonstrate a statistically significant relationship with final grade. As was noted in the evidence for rejecting null Hypothesis 4a, the instructor block often results in a proportionately larger increase in R-square or amount of variance explained. This was found for each level in English, after all other blocks of variables have been added to the model.

For college level English courses, placement test scores, demographic, and dispositional variables account for approximately 32% of the variance in final grade. Situational variables add a small, but still statistically significant, amount of explanatory power. Instructor variables explain approximately 13% of the variance in final grade. This is shown by the accompanying F test to be highly significant statistically. The proportion of variance explained by instructor effects is thus greater than the proportion explained by placement test scores. This is further evidence of the plausibility of Alternate Hypothesis 4a.

For English courses one level below college, placement test scores, demographic, and dispositional variables account for approximately 22% of the variance in final grade, while the addition of the situational block of variables improves R-square by approximately 2%. Instructor variables reduce a significant amount of error in the model. Approximately 18% of the variance in final grade is accounted for by instructor variation. This is also greater than the proportion of variance accounted for by placement test scores.

For courses two levels below college level, the explanatory model appears to explain a somewhat greater amount of variation than for the two higher levels. After all categories of variables are entered into model, R-square increases to approximately .50. This compares to final R-squares of .42 and .48 for college level and courses one level below college, respectively. For non-credit courses, including the instructor in the model explains a statistically significant proportion of variance in final grade.

Mathematics Courses

With the exception of college level mathematics courses, placement test scores, demographic, dispositional, and situational variables were statistically significant with respect to explaining variance in course grade. For the three levels of mathematics, instructor variables were significant in explaining variance in final grade.

College level mathematics courses showed the greatest departure from the general pattern of relationships between the independent and dependent variables noted in the other two levels. For example, in College level mathematics placement test scores and demographic variables were not found to have a statistically significant relationship with final grade. However dispositional variables accounted for a proportionately greater amount of variance in final grade than other variables in the model. At lower levels in the curriculum, dispositional variables also contribute a significant amount of explanatory power to the course grade model. The entry of instructor codes into the model show a statistically significant relationship with the dependent variable of final grade for all three levels of mathematics. In all cases the proportion of variance in final grade explained by instructor characteristics is greater than the proportion accounted for by mathematics placement test scores.

Table 5.29: Course Performance Model; Final Course Grade in Mathematics by Curriculum Level

Level	Variables	Multiple R	R Squared	F	Significance of Change
College					
	Placement Test Scores	.06	.003	.651	p>.400
	Demographic	.33	.11	1.96	p<.100
	Dispositional	.58	.34	2.60	p<.000
	Situational	.63	.40	2.51	p<.000
	Instructor ID	.77	.59	1.99	p<.000

Level	Variables	Multiple R	R Squared	F	Significance of Change
One Level Below (pre-college)	Placement Test Scores	.14	.02	10.08	p<.005
	Demographic	.42	.17	8.92	p<.000
	Dispositional	.48	.23	4.14	p<.000
	Situational	.53	.28	4.15	p<.000
	Instructor ID	.68	.46	2.64	p<.000
Two Levels Below (basic skills)	Placement Test Scores	.18	.03	12.19	p<.002
	Demographic	.33	.10	3.94	p<.000
	Dispositional	.40	.16	1.90	p<.004
	Situational	.45	.20	2.00	p<.000
	Instructor ID	.61	.38	2.27	p<.000

Retention in English Courses

For college level English courses, placement test scores, dispositional, and instructor variables showed a statistically significant relationship with the dependent variable of retention. However this same pattern was not found for the other two levels of English analyzed. Placement test scores and dispositional variables were not found to be significantly related to retention for one and two levels below college English courses. At the lowest levels of English (two levels below college) situational variables contributed significantly to improving the prediction of the retention criterion. For courses below college level, demographic variables showed a stronger association with retention than other categories of variables used in this model. With the exception of college level English, instructor variables were not found to be significantly associated with the probability of retention in the course. Thus there is evidence that instructor effects are not as significant in predicting retention in the lower English course levels. At the lowest level of English, demographic and situational variables significantly contribute to improvement in the -2 log-likelihood of the retention model.

Table 5.30: Retention Model for English; By Curriculum Level

Level	Variables	Model Chi-Square	-2 Log-Likelihood Change	Significance Chi-Square Improvement
College	Placement Test Score	11.3	571.2	p<.004
	Demographic	8.8	562.4	NS
	Dispositional	47.0	515.4	p<.002
	Situational	8.1	507.3	NS
	Instructor ID	60.0	447.3	p<.05
One Level Below (pre-college)	Placement Test Score	2.3	848.9	NS
	Demographic	24.2	824.7	p<.008
	Dispositional	25.1	799.6	NS
	Situational	10.1	789.4	NS
	Instructor ID	65.8	723.6	NS
Two Levels Below (basic skills)	Placement Test Score	5.6	225.5	NS
	Demographic	25.8	199.6	p<.004
	Dispositional	29.2	170.5	NS
	Situational	23.4	147.0	p<.003
	Instructor ID	25.8	121.2	NS

Retention in Mathematics Courses

In college level mathematics, including the instructor added significantly to the explanatory power of the retention model. For courses one level below college level, instructor effects were not found to be significant, while for courses three levels below college level, instructor effects were significant. Other significant variables included dispositional variables for college level and courses one level below college. Dispositional variables were not found to be significantly related to retention in courses three levels

below college. The categories of variables that showed the strongest association with retention for courses two levels below college level were placement test score and instructor variables. Demographic and situational variables were not found to be significantly associated with retention at courses two levels below college level mathematics.

Table 5.31: Explaining Variance in Mathematics Retention; by Curriculum Level

Level	Variables	Model Chi-Square	-2 Log-Likelihood Change	Significance Chi-Square Improvement
College	Placement Test Score	3.2	716.0	p<.08
	Demographic	17.9	698.1	p<.06
	Dispositional	47.0	650.4	p<.002
	Situational	7.9	642.4	NS
	Instructor ID	104.5	537.9	p<.04
One Level Below (pre-college)	Placement Test Score	.08	389.7	NS
	Demographic	13.6	376.1	NS
	Dispositional	43.8	332.4	p<.003
	Situational	12.2	320.2	NS
	Instructor ID	50.1	270.0	NS
Two Levels Below (basic skills)	Placement Test Score	6.4	413.0	p<.02
	Demographic	15.5	397.6	NS
	Dispositional	31.0	366.5	p<.10
	Situational	6.45	360.0	NS
	Instructor ID	70.3	289.8	p<.000

The null Hypothesis 4a posited that instructor characteristics will not account for more variance in final grade and retention than will placement test scores. This hypothesis was tentatively rejected for English and mathematics courses with respect to final grade.

However, the effect of the instructor on student retention was less obvious. With the exception of college level English and mathematics, and non-credit mathematics courses, instructor effects did not add significant predictive power to the retention model. When analyzed for all levels of English and mathematics, instructor characteristics were found significantly related to retention, however, this may be a function of a larger sample size. At the intermediate levels in the curriculum, including the instructor in the model did improve the prediction, but the improvement was not statistically significant. This finding merits further research. Nonetheless, the effects of the instructor on retention was sufficiently evident to tentatively accept Alternate Hypothesis 4a.

Instructor Employment Status and Student Outcomes

The Null Hypothesis 4b posited that:

> There is no relationship between instructor status (full- or part-time employment) and student final grades and course retention after controlling for student characteristics and test scores.

To test Hypothesis 4b, several analyses were conducted. The first analysis compared the test score, GPA1, GPA2, and retention, of study participants by instructor employment status. The comparison was done to determine if the performance of students appeared to differ significantly by instructor employment status. The data shown in table 5.32 suggest that the grades, test scores, and retention outcomes of study participants in English courses differed only slightly between the two groups of instructors. Although the differences in grade point averages and writing test scores were statistically significant, the differences in practical terms appear minimal. There does not appear to be a great difference either in the entry level abilities of students or in the outcomes of final grade and retention when analyzed by instructor employment status.

Table 5.32: Comparison of Test Scores, GPA1, GPA2, and Retention, Between Full- and Part-Time English Instructors

Employee Status	Full-Time	Part-Time	Significance
GPA1	2.3	2.4	p<.02*
GPA2	1.8	1.9	p<.03*
Proportion Retained	78	79	NS**
Reading Test	17.9	17.9	NS*
Writing Test	21.9	21.3	p<.003*

* significance based on t statistic ** significance based on phi statistic

When analyzed for mathematics courses, significant differences were found in both the entering abilities of students (mathematics placement test score) and the outcomes (grade point average) between full- and part-time instructors (table 5.33). The average grade awarded by full-time instructors was slightly higher than those awarded by part-time instructors. However, full-time instructors tended to have students with higher entering mathematics ability as measured by test score. There was no significant difference found in the retention rate of full-time compared with part-time instructors.

Table 5.33: Comparison of Test Scores, GPA1, GPA2, and Retention, Between Full- and Part-Time Mathematics Instructors

Employee Status	Full-Time	Part-Time	Significance
GPA1	2.8	2.6	p<.000*
GPA2	2.3	2.0	p<.000*
Proportion Retained	80	78	NS**
Mathematics Test	26.5	22.6	p<.000*

* significance based on t statistic ** significance based on phi statistic

To determine grading differences between instructors within the two employment categories, a one-way ANOVA for independent data was performed using student grades as the criterion in table 5.34. The resulting F values were statistically significant and the resulting eta-squared (the sum of squares between instructors divided by the total sum of squares) indicates that from approximately 10% to 19% of the variation in student grades can be attributed to instructor differences. There appears to be somewhat greater variation

among part-time instructors than full-time instructors in both English and mathematics. This analysis provided evidence of a possible difference between the two groups of instructors in grading variation. This led to the calculation of separate regressions for full-time and part-time instructors to determine if grading variation was related to the employment status of the instructor.

Table 5.34: Analysis of Variance of Grade Point Average Within Instructor Categories For English and Mathematics

Subject Instructor Status	F Statistic	Significance	Eta Squared
English			
Full-Time	4.40	p<.000	.137
Part-Time	4.61	p<.000	.194
Mathematics			
Full-Time	2.50	p<.000	.098
Part-Time	3.30	p<.000	.185

Tables 5.35 and 5.36 present the multiple regression results for both full- and part-time instructors in English. These data tend to confirm the results of the one-way ANOVA with part-time instructors demonstrating greater variation in grading than the full-time instructors. This was supported by the proportionately greater amount of variance explained by part-time instructor identification codes compared with full-time instructor identification codes.

Table 5.35: Factors Explaining Variance in English Course Grade by Full-Time and Part-Time Instructors

Block	Multiple R	R Squared	F	Significance of Change
Full-Time Instructors				
Placement				
Test Scores	.26	.065	21.2	p<.000
Demographic	.35	.122	6.9	p<.000
Dispositional	.43	.183	3.8	p<.000
Situational	.45	.206	3.4	p<.000
Instructor ID	.56	.312	2.7	p<.000
Part-Time Instructors				
Placement				
Test Scores	.24	.057	15.5	p<.000
Demographic	.39	.154	7.5	p<.000
Dispositional	.51	.264	5.0	p<.000
Situational	.52	.275	4.1	p<.000
Instructor ID	.68	.462	3.7	p<.000

With respect to mathematics courses, there were also some differences between full- and part-time instructors in the amount of variance explained in final grade. As was found in explaining variance in final grade for English courses, part-time instructors tended to show greater grading variation than full-time instructors. In explaining variance in final grade, dispositional variables appear to explain a significant amount of variance. Situational variables also contribute significantly to the regression equation.

Table 5.36: Explaining Variance in Course Grade in Mathematics Courses by Full-Time and Part-Time Instructors

Block	Multiple R	R Squared	F	Significance of Change
Full-Time Instructors				
Placement				
Test Scores	.14	.020	3.55	p<.07
Demographic	.47	.221	4.09	p<.000
Dispositional	.63	.397	2.71	p<.000
Situational	.65	.423	2.36	p<.000
Instructor ID	.77	.596	2.31	p<.000
Part-Time Instructors				
Placement				
Test Scores	.19	.037	9.66	p<.003
Demographic	.42	.177	4.73	p<.000
Dispositional	.54	.296	2.80	p<.000
Situational	.61	.374	3.07	p<.000
Instructor ID	.76	.577	2.78	p<.000

Retention by Full- and Part-Time Instructor

Retention is somewhat dependent on the employment status of the instructor. The reduction in the -2 log-likelihood is greater when part-time instructors are entered into the retention model compared with the contribution of full-time instructors to the model. However, including either group of instructors in the retention model improves the explanatory power of the model significantly. This is evidenced by the significant change in the -2 log-likelihood for both full-time and part-time instructors in predicting the retention criterion.

Table 5.37: Explaining Variance for Retention in English Courses by Full- and Part-Time Instructors

Block	Model Chi-Square	-2 Log-Likelihood Change	Significance Chi-Square Improvement
Full-Time Instructors			
Placement			
Test Score	13.0	851.9	p<.002
Demographic	30.9	821.0	p<.000
Dispositional	35.2	785.7	p<.05
Situational	12.8	772.9	NS
Instructor ID	61.7	711.1	p<.05
Part-Time Instructors			
Placement			
Test Score	4.02	707.4	NS
Demographic	15.7	691.7	NS
Dispositional	31.2	660.6	p<.10
Situational	15.7	644.8	p<.05
Instructor ID	75.7	569.0	p<.003

For mathematics courses, the effect of instructor employment status was significant only for part-time instructors.. Among full-time instructors, placement test score, and dispositional variables contributed to a significant reduction in the -2 log-likelihood of the model. Among part-time instructors, demographic and instructor variables were found to contribute significantly to reduction in the -2 log-likelihood of the retention model.

Table 5.38: All Mathematics Courses; Explaining Variance for Retention in Mathematics Courses by Full- and Part-Time Instructors

Block	Model Chi-Square	-2 Log-Likelihood Change	Significance Chi-Square Improvement
Full-Time Instructors			
Placement			
Test Score	3.20	264.7	p<.08
Demographic	5.33	259.4	NS
Dispositional	50.57	208.8	p<.000
Situational	5.80	203.0	NS
Instructor ID	23.53	179.4	NS
Part-Time Instructors			
Placement			
Test Score	1.03	353.7	NS
Demographic	20.98	332.7	p<.03
Dispositional	29.20	303.5	NS
Situational	4.78	298.7	NS
Instructor ID	82.74	216.0	p<.000

Instructor Employment Status and Student Performance-Summary

Part IV of the analysis compared the contribution of full- and part-time instructors to the course performance outcomes of study participants. The evidence provided was sufficient to reject the null Hypothesis 4b that posited no relation between the employment status of the instructor and student course performance outcomes. Although both groups of instructors explain significant amounts of variance in final grade, part-time instructors explain a proportionately greater amount of variance in final grade than do full-time instructors. This evidence led to the acceptance of Alternate Hypothesis 4b when final grade was the criterion.

With respect to predicting the criterion of retention, full-time instructors in mathematics did not contribute significantly to the logistic regression model as evidenced by the small reduction in the -2 log-likelihood of the model (table 5.38). However, for part-time instructors, including the instructor as part of the model did improve the

predictive power of the model significantly ($p < .000$). For English courses, both full- and part-time instructors added significantly to the predictive power of the logistic model, however, part-time instructors tended to improve the retention model more substantially than full-time instructors. This suggests greater inconsistency among part-time instructors in the retention outcomes of their students, after the effects of other variables have been controlled statistically. This evidence of greater variation in retention outcomes between full- and part-time instructors was sufficient to accept Hypothesis 4b. The employment status of the instructor does appear to make a difference in both the final grade and retention outcomes of study participants.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

The present investigation has explored the multiple sources and effects of variance on college course performance outcomes. This chapter will report the findings relevant to each primary research question guiding the investigation. The final section of this chapter focuses on practical and theoretical issues and provides recommendations for future research.

The research questions and the hypotheses derived from them were guided by an analysis of current social and public policy issues with respect to predictive validity and practices of placement testing in the community colleges. A review of the historical development and application of placement testing in educational and training settings provided a useful context to study the present problem of predictive validity in the community colleges. The first question asked:

> Are placement tests highly predictive of course performance outcomes such as course grades or retention in college courses?

Placement tests are increasingly used to allocate access to college courses, employment, the military, and high demand training programs. The increased use of placement testing to allocate opportunity and access has recently come under the attention of public policymakers at the state and federal levels. The use of placement or entry tests to determine who can profit from instruction or has aptitude for a particular profession has been an administratively convenient method to allocate opportunity. In the community colleges, particularly since the passage of Matriculation reform in the late 1980's, the use of standardized placement tests in the community colleges has shown a remarkable growth (Alkin and Freeman, 1991). Most placement tests are used to place students at a particular

level in the curriculum. Others are used to allocate openings to applicants for high demand occupational programs.

However, a host of regulations and mandates has proliferated concurrent with the growth in the use of standardized placement testing in the community colleges. Public policy mandates in the area of testing have required community colleges and other institutions to conduct extensive studies to demonstrate the predictive validity of placement tests used to screen and classify students. Influenced by court cases and other litigation, California state education code requires that colleges demonstrate that the predictive validity coefficient between a placement test score and final grade achieve at least a .35 level.

As suggested by this study, achieving a mandated coefficient of .35 is difficult. There are several technical constraints that limit the size of the predictive validity coefficient. These include the restriction of range due to the evaluation of an already existing placement system, the lack of common scaling between final grade and a placement test, instability in the placement test, and instability in the criterion.

These testing regulations have served to spawn a growing cottage testing industry in the state's community colleges. The then incipient practice of using placement tests in the community college did not seem wise, even years before the current debates given the lack of institutional experience and a dearth of capable psychometricians at the community colleges to fully understand the capabilities and limitations of testing (Sheldon, 1970). Soon after the state mandated placement testing in the community colleges, MALDEF filed a lawsuit charging several colleges with using cutting scores on placement tests to unfairly deny students access to collegiate-level courses that they could have passed. Immediately after the lawsuit was filed, the state promulgated a vast array of testing guidelines. The colleges were then compelled to seek expertise in the form of consultants,

psychometricians, and educational researchers to comply with the complex set of rules and mandates regarding the use of standardized placement tests.

The mandate for establishing predictive validity resulted in a raft of predictive validity studies in the community colleges. A review of predictive validity studies and of relevant testing and measurement literature suggested that the relationship between placement test scores and final grades or retention failed to meet the state required .35 coefficient for test validation.

To determine the relation of placement test scores and the criterion variables of course grade and retention, predictive validity coefficients were obtained for three levels of English and mathematics. To reduce the threat of artificially lowered coefficients due to attenuation of range in the independent variable of test score, each coefficient was corrected for a restriction of range for each level of English and mathematics. Evidence was obtained for the predictive validity of a placement test to another placement test using two administrations of a reading and writing placement test to a cohort of students, the experiment used to test Hypothesis 1b.

For all levels of English and mathematics analyzed, placement test scores were found to have a weak to modest, yet statistically significant relationship ($p<.05$) to the criterion variables of course grade and retention. Corrected for restriction of range, improved the strength of the coefficients; however, only in the case of college level English courses did the coefficient exceed the .35 threshold mandated by state law.

Evidence was accept to accept Hypothesis 1b was gathered using the correlation coefficients between the two test scores. The correlation coefficients revealed that elements in common with predictor and criterion yielded greater predictive validity than between test score and final grade.

The relatively weak relationship of placement test scores to course outcomes was further confirmed by evidence showing that even after a semester of instruction, student test scores from one administration to the next showed no significant differences. A fundamental question is raised then as to the purposes of pre-enrollment placement testing that displays little relationship to course performance but a strong relationship to another test score several months later. Instruction did not markedly improve placement test scores over time. Misalignment of the objectives for success in a course with the skills demanded by standardized placement testing may be responsible for the insignificant change in test scores over the two test administrations. Minimal change in test scores was also noted when students had completed a level of English and enrolled in a higher level English course the following term and achieved virtually the same scores on the reading and writing placement tests.

The Relation of Student Course Eligibility to Course Success

A comparison of the outcomes of students considered ineligible for a course with the outcomes of students considered eligible suggested little practical difference. Course eligibility was determined either through a score-based inference (cutting score on a placement test) or by completion of a course considered to be a prerequisite for the course. During a period when ineligible students could ignore enrollment advice and attempt any course they chose, a comparison of the outcomes of those eligible with the ineligible was possible. A comparison of outcomes for seven subject areas found that although there was a statistically reliable association between eligibility and course success, the effect was weak to modest. The success rates of the ineligible once computed as a proportion of the success rates for the eligible students confirmed the minimal practical effect of a prerequisite on course outcome. In fact, ineligible students who attempted the course were

generally 70%-80% as successful as the eligible students. It was therefore difficult to conclude that the ineligible students are indeed highly unlikely to succeed. The implementation of mandatory prerequisites is contingent upon such evidence; however, when successful students are classified as probable failures, access is negatively affected. The cost to the student in terms of opportunity costs, and to the institution due to inefficient allocation of instructional resources are probably considerable. Under a mandatory system of placement, analysts may never know how many students are misclassified as failures. By design, mandatory placement systems prevent ineligible students from attempting a course without going through and extensive appeal and challenge process. The present study suggests, as did the studies conducted in the military services a quarter century ago, that many individuals deemed as probable failures can in fact succeed if given the opportunity.

There are other methods of determining readiness for a course. Student interviews with the instructor familiar with the requirements of the course may help determine whether the student has sufficient communication or computation skills to attempt the course. A review of the students' contextual grade point average in college courses may also be instructive. For example, if a student has completed twelve credit units with a grade of 'C' or better, there is good evidence that the student has the motivation and skills to succeed in many college courses. In the case of vocational courses, the life experience or employment history of the individuals may provide evidence that they have the prerequisite skills to take a photography course or a course in office information systems, even if they failed to achieve minimum cutting scores on a mathematics or English placement test. Although this expanded method of ascertaining eligibility may require more effort and cost, it may serve students better by more accurately classifying them as potentially successful, rather than highly unlikely to succeed.

An expanded appraisal of student abilities and experiences would possibly reap additional benefits. Discussions and interviews with faculty and counselors may also help the student to develop more realistic expectations of what is required of them in college courses. As part of the Matriculation program, students are to become more of a partner in achieving their educational goals rather than passive recipients of instruction and services. Engaging the student in dialogue with counselors or faculty regarding their expectations and abilities may help the student to make more reasoned choices. Many students may decide they do in fact need some remediation in English or mathematics to achieve their goal of a degree, employment, or transfer. Greater engagement of the student with the college may improve outcomes for the learner and the institution.

The findings of a weak relationship between test scores and the dependent variables of final grade and retention, although anticipated, militate for colleges and other institutional to engage in closer scrutiny of the value of such score-based inferences to allocate opportunity for education and training. State regulations to mandate minimum predictive validity coefficients do little to change the repeatedly observed findings that there are few common elements among placement test scores and course grades and retention. Mandating a predictive validity coefficient without attention to the objectives of the course or skills needed for retention, greatly increases the likelihood that students, many of whom are capable of passing, will be misclassified as failures.

This investigation could be conducted in other educational and training settings that rely on test scores for admittance, but use supervisor or teacher ratings for advancement. The lack of correspondence between the predictor and criterion observed in this investigation may also exist in other educational, employment, or training organizations. As was noted in the literature review chapter, a comprehensive review of the military experience in accidentally accessing hundreds of thousands of ineligible recruits who

subsequently entered military service revealed that when the criterion for advancement was supervisor ratings, these recruits performed almost as well the higher aptitude, eligible recruits.

The results from this investigation may help to promote faculty development among academic departments concerned with the goals and objectives of courses and how the goals and objectives might be more closely aligned with assessment practices. Such an alignment may help academic or vocational faculty to reach more common agreement on the standards for success and promotion. Establishing standards might be particularly helpful in the community colleges, where much of the instruction is conducted by part-time faculty, often at off-campus locations. Including the part-time faculty in discussions about standards and the goals of placement testing may help reduce the inconsistency in grading practices. This has implications for faculty development in recognition of the necessity to align courses. because faculty teach these courses and accountable for the learning outcomes of students. It might be in the best interest of the college to consider the consider the alignment of placement tests with the objectives of the course.

The second research question asked:

> How do student characteristics affect the prediction of performance outcomes?

An important objective of this investigation was to gain a better understanding of the relationships of test scores and student biographical data with course outcomes to help guide the development of college course placement models in future research and application.

Part III of the study was informed by both public policy and testing theory regarding the use of multiple or companion measures in addition to test scores for use in student placement in the community colleges. This part of the study was also informed by

the tenets of Point-to-Point theory that suggested that the prior activities and achievement of the student would show a stronger relationship to course grades and retention than test scores.

The main objective of this part of the investigation was to develop a composite model that would explain a proportionally greater amount of variance in the course performance outcomes of students than test scores alone. The inclusion of student demographic, situational, and dispositional data in a model explaining variance in course grades or predicting retention added greatly to the explanatory power of the models. For both English and mathematics courses, student characteristics, traits, and activities showed a stronger relationship to the course outcomes than did placement test scores. Of particular explanatory power were those variables that reflected the prior academic performance and dispositional characteristics of the student. Dispositional variables such as high school grade point average, grades earned in prior English or mathematics courses, and the importance of college to the student were found to have a statistically significant relationship with the dependent variables of final grade and retention. The proportion of variance explained in the dependent variables by dispositional variables generally exceeded the variance explained by test score, demographic, or situational variables.

In the literature review conducted for this investigation, it was noted that almost thirty years ago, Sheldon (1970) wondered if accurate placement information about a student could more easily be gleaned from a review of high school transcripts or even self-reported high school grade point average than from mass administrations of standardized placement tests. Sheldon's supposition that student dispositional information is of more predictive value than placement tests in the community colleges was confirmed in this investigation. An important ancillary objective of the investigation was to understand the

value of identifying and including other relevant student data in a model that more accurately diagnosed and forecast student performance.

Many of the data available to enhance the prediction of student success might be used by community college staff for student placement, but are often left unanalyzed and unused. These findings could be especially salient to colleges whose placement practices rely heavily on the use of standardized tests. The approach of this investigation could be particularly helpful to institutions wishing to conduct predictive validity studies. The method described here may improve the inferences about student aptitude and hence the accuracy and fairness of the prediction of final grade or to enhance retention. In some cases, it may be preferable to eschew the use of standardized placement tests for student placement and instead rely on dispositional data to place students at the level in the curriculum that bests meets their instructional needs.

The present study repeatedly found that with few exceptions, dispositional variables had more explanatory power than situational variables. Cross (1981), however, in her review of the reasons why students drop out of school, tended to find that situational variables were often the reason. The contrary findings of the present investigation may be the result of relatively few situational variables gathered by the questionnaire. The situational variables that were gathered perhaps were not the salient variables to accurately describe the situation of the student. For predictive validity purposes, situational variables may be of less value because of their relative volatility. That is, almost by definition situational variables are affected by the vicissitudes of a student's life, family, and employment responsibilities. Dispositional variables may be of greater predictive power for retention and final grade because they may reflect the more enduring characteristics of the student that portend their likelihood for success.

An anticipated outcome of this investigation was to identify methods for colleges or other education and training institutions to improve the validity of their instruments and inferences they make about individual student aptitude and ability. Improving predictive validity is particularly relevant for the open-access community college because the open admissions policies of the community college assure that matriculating students will display a wide range of academic ability. Given the limited information available from standardized placement tests, it is likely that certain dispositional, situational, or behavioral traits are more useful for predicting course success. Colleges may wish to use such findings to their advantage in placing students.

The composite models for final grade and retention also suggest that test scores do not work as well when used as the sole basis for selecting and classifying students. The dispositional and situational variables that best explain retention or success in a course are not measured well nor gathered by standardized assessment tests. If they are gathered, institutions often are not able to integrate the biographical information about the student in a way that enhances the likelihood for retention or course success. It is hoped that this study will stimulate discussion of how dispositional variables may play a greater role in the assessment and placement of students.

Research question 3 asked:

How do instructor characteristics affect the prediction of performance outcomes?

Perhaps the most important finding with respect to research question three was that accurate and equitable placement is contingent on stability in the criterion. When misclassification depends on the course one enrolls in, then the placement process is hindered, and the objective of improving student outcomes is confounded.

The quantification of instructor variance in the prediction of final grade and retention is instructive to the development of recommendations to better understand and

improve the predicting of student performance or allocating opportunity based on standardized test scores. Such recommendations could be used by college faculty, administrators, measurement specialists, and state policymakers to better link the objectives of placement testing with the objectives of instruction.

Another interesting finding was that for many of the courses analyzed, instructor grading variation accounted for more variance in the grade prediction model than placement test scores.

Instructional treatments or approaches may be vastly different as evidenced by different instructor characteristics and practices of teaching and evaluation. As suggested by Willingham (1974), the interaction of individual aptitude and propensity toward retention with the treatment or instructional methods of the faculty member contribute strongly to instructional outcomes. Willingham's observations were generally confirmed in this study.

Theory was also instructive in this phase of the investigation. Aptitude-Treatment Interaction theory led to the hypothesis that the instructional treatment and outcome criteria would be highly variable, thus including the instructor as a source of variance would explain significant error in the model. For this part of the investigation, the objective was to apply principles derived from aptitude-treatment interaction theory to community college student placement practices. The application and modeling of this theory provided strong evidence of significant instructor effects on the outcomes of final grade and retention. This was done through the use of instructor codes entered as dummy variables into a linear regression model for explaining final grade, and a logistic regression model predicting the probability of retention.

Alternate Hypothesis 4b focused on the effect of the employment status of the instructor on final grade and retention. One of the underlying assumptions of community college reform legislation was that instructional quality in the state's two year colleges

would improve by increasing the ratio of full-time instructors to part-time instructors. Assembly Bill 1725 in fact mandated that colleges strive to have seventy five percent of their courses taught by full-time professional faculty. The architects of this reform believed that full-time faculty would have more identity and connection with the college and that this connection would enhance the instructional standards of the community colleges.

Using Alternate Hypothesis 4b, this study investigated if the variance in the criterion variables was greater for part-time versus full-time instructors. ANOVA results suggested greater variance in grading by part-time instructors, while retention rates did not differ significantly between the two groups of instructors. Although both groups of instructors contributed significantly to variance in the criterion, separate regressions for full- and part-time instructors confirmed greater inconsistency in grading among the part-time faculty. This finding tends to support the belief that a higher proportion of full-time instructors in the college will perhaps better maintain the instructional objectives and standards of courses. There may be reasons for this finding. For example, in discussions with community college faculty and staff, many have reasoned that part- time instructors will have higher retention rates because of their need to maintain a minimum class size so that the class is not canceled. On the other hand, others suggest that the retention rates of full- time instructors are higher because they are more connected to the college, have offices for meeting students, attend more faculty meetings, and have more consistent practices with respect to evaluating student work. This is consistent with the philosophical underpinnings of community college reform legislation mandating a higher proportion of full-time instructors in the community colleges.

The most interesting assumption I encountered in my research and discussions with college faculty and staff on the issue of placement testing was the belief that variance exists solely within the students. The philosophy and practice behind placement programs

typically assume that the instructional treatments are equivalent and the criteria for success are reliable within treatments (courses). Therefore it appears likely that the implementation of a placement system that mandates placement at certain curricular levels or excludes access to programs through the use of cutting scores or prerequisite courses may lack precision in predicting student outcomes reliably. Within certain academic departments, there appears to be a largely uncritical acceptance of the information obtained from standardized placement tests. To some, placement tests can place students accurately in the curriculum, prevent others from enrolling in courses where they are likely to fail, enhance retention, and diagnose educational deficiencies.

As suggested by Aptitude-Treatment Interaction and I believe confirmed by this study, it is the interaction of aptitude and treatment that produces learning and retention outcomes. What is needed is a better understanding of the nature of that interaction and how it differs with individual aptitude. Placement can be improved when other potentially relevant dispositional and situational information about the student or instructional approach of the faculty member is gathered and analyzed. This relevant information is fortunately often available at the college and may be used to enhance the interaction of individual abilities and interest with instructional approaches.

The introductory chapter discussed the usefulness of ability grouping as a strategy to improve collegiate education. There is a belief from the elementary schools to the community colleges that ability grouping results in better instructional outcomes. While this belief is disputed (Oakes, 1985; Tyler, 1974), it is not clear if ability groups can be reliably created using standardized placement tests. As suggested earlier, placement tests sample only a small portion of what a person knows or is able to do. The demographic, dispositional, and situational profile of the students included in this investigation suggest that even when score ranges are used to allocate access to certain courses, the groups do

not appear homogeneous. There is still a tremendous diversity in background, school experiences, traits, beliefs in self-efficacy, and attribution. Setting score ranges on a standardized test for entry into a class may serve to restrict the range of ability only in some small way. However evidence suggests that dispositional variation is the norm rather than the exception in students supposedly grouped by ability. The notion of ability grouping is also confounded when the instability of the criterion of final grade or retention is high. What may matter more is the instructor a student chooses.

There are implications here for retention theory. As described in the literature review, many retention theorists attribute higher retention with greater amounts of student involvement and contact with the institution. For example, student involvement theory (Astin, 1985) posits that enhanced retention is largely an outcome of greater student involvement in the institution. Findings that support this theory or variants on this general theme of involvement have been reported by several authors (Pascarella & Terenzini, 1991; Bean, 1982). The evidence presented in this investigation may serve to augment student involvement theory by more closely focusing on the nature and quality of student involvement. This investigation suggests that certain instructors contribute a significant amount toward predicting the course success or retention of students. The question then becomes not one of degrees of involvement, or the amount of psychological and physical energy one invests in college, but perhaps rather the aspects and quality of the actual interaction between students and faculty. Given that most community college students commute to college and often spend little time on campus, being selective about whom the student chooses to spend time with on campus, particularly in the classroom, appears to matter.

Outside of the community colleges or higher education, the findings of this investigation may be helpful to employers who face potential legal challenges to promotion

or application procedures that rely on employment tests. The recommendations may be relevant to employers who use a test for job entry, but promote individuals on the basis of supervisor ratings. An example of this practice can be found in the armed services where eligibility for certain higher status training and occupations is allocated for many recruits by test scores yet promotion is determined by supervisor ratings of effectiveness and aptitude. The model developed by this investigation may reduce the potential problem of a student or employee's being misclassified as a failure when in fact their dispositional characteristics suggest strongly that they could succeed if given the opportunity.

The salience of the findings of multiple sources of variance may be particularly useful in that the decisions made about an individual's eligibility for a class or program of study are often considered "high stakes" decisions. That is, the inferences made about an individual from a test score may have a strong impact on their economic, social, or psychological, well-being. For example, a person denied entrance into an occupational program on the basis of a cutting score on a test may be dramatically affected in their career choices and potential lifetime earnings. Or an individual placed in non-credit courses in college not only has opportunity costs associated with prolonging their college career, but also may become ineligible for federal financial aid. Under federal Ability-to-Benefit legislation, students failing to achieve a minimum cutting score on certain nationally standardized tests are deemed unable to "profit from instruction" and as such are ineligible for immediate consideration for student financial aid. Thus such test score-based inferences may be viewed as a high stakes decision for the individual when they are used to allocate access to courses or programs of study.

Misclassification is thus more than an incorrect inference derived from a score. This investigation has served to highlight the importance of the criterion for success. Misclassification may primarily be a function of who is assigning the grade. Poor

prediction of performance or misclassification is exacerbated when the criterion for student success can vary depending on the class one enrolls in. Specifying strict achievement test scores may mean little in improving student outcomes if the criterion for success is not based on clear, reliable, behavioral objectives for success in class. Identification of the extent of instructor effects on course grade and retention may assist college academic departments in focusing more closely on course objectives and the consistency in application of these objectives for course content. This might be particularly useful for departments considering the use of standardized placement tests for entry or for placement. This investigation suggests that it is difficult to have it both ways. From an equity standpoint, one cannot specify strict entry standards based on rigidly enforced test rules and cutting scores, yet have a criterion for success that is highly unreliable. When ineligible students are 70-80 percent as successful as eligible students when given a chance to enroll in a course they would otherwise be prevented from enrolling in, the criterion may bear closer scrutiny.

Another recommendation would be for colleges to more clearly recognize the principles of Point-to-Point theory and Aptitude-Treatment Interaction to develop and implement more sensitive and relevant measures of student aptitude. Identifying potentially useful measures of student aptitude and characteristics and matching these measures to instructional treatments may produce more reliable outcomes for students. It is difficult and often unfair to argue that a rigid course placement system needs to be implemented to group students by ability if instructor grading variation contributes a significant amount of error to the final course outcome. It is also inequitable to designate many students as ineligible for a class because they are highly unlikely to succeed, yet when they are allowed to enroll despite this advice, they are approximately 70 to 80 percent as successful as the eligible group. This study not only confirmed the existence of substantial grading and retention

variation that confounds prediction of success, but also attempted to quantify the amount as a proportion of the final grade or retention.

This study also sheds more light on the often repeated, but seldom understood notion of academic "standards." In reviewing the founding documents for the Matriculation legislation it was noted by some members of the State Academic Senate that academic standards in the California Community Colleges were declining. This decline in standards were linked to diminished transfer rates, graduation rates, and lower productivity of vocational graduates. A review of Matriculation documents suggested that academic standards are represented by the selectivity of the institution, that is, the caliber of the entering students are primarily responsible for the academic standards of the institution. This investigation suggests that academic standards might best be thought of what the institution imposes on students at the exit point, not what the students bring with them at entry. Academic standards are brought to bear on the evaluation of student performance and are institutionally derived and defined. To equate declining academic standards with the lower ability of entering students in the community colleges I believe is a misrepresentation of the notion. As students enter the college from increasingly diverse backgrounds and heritage, the imposition of consistent academic standards is the duty of the academic and policy leaders in the colleges.

The view that the academic standards of the institution reside in the student body is not uncommon even in open-access institutions where selective admissions is not practiced by statute. The Academic Senate for the California Community Colleges as part of the resolution endorsing mandatory placement testing for students seems to also have equated academic standards with the educational background and traits of incoming students. They maintained that declining academic standards in turn precipitated a decline in the quality of programs and reputation of the state's two year colleges (Alkin and Freeman, 1991).

However Cohen and Brawer (1989) observed that although serving underprepared students is an essential element of the two-year college mission, serving increasing numbers of underprepared students should not translate into lowered academic standards.

There are other possible lessons to be learned from this investigation. One would involve a review of the goals of placement testing and the goals of an instructional sequence. More closely aligning a college's assessment and placement program with the desired behavioral objectives of a course would probably improve predictive validity and also improve the instructional outcomes for students (Cohen, 1970; Dubin and Taveggia, 1968). Matching assessment methods with curricular content would help students make more informed choices about their own academic abilities and aptitude. To the extent that college assessment practices are dissimilar to the actual skills needed for course success, then the assessment and placement goal of creating homogeneous student groupings may remain elusive.

The problem of student placement and predictive validity in the community colleges is complex. As noted by Cohen and Brawer (1987) in their analysis of the use of placement test validity in the community colleges:

> Obtaining higher correlations is difficult, because the variables are inconstant. The psychometrists ask that the dependent variable, the course grades, be more reliable; the faculty seek tests that will predict student success regardless of the shifting criteria for grades. The result is that the faculty prefer to rely on their own measures, particularly of student writing skills. For obvious reasons, writing assessment is considered a better predictor of student grades in the English classes; the same staff who are marking the writing samples on the entrance examination are marking the writing assignments in the classes. The writing sample as an entrance test, then, is a behavior equivalent to that expected of students in class, while performance on a quick score test of word usage is at most analogous behavior.

Implicitly recognizing the constructs of Point-to-Point theory, Cohen also notes that

> Some of the more astute institutional researchers and faculty members have recommended using assessment tests that more closely approximate the behavior demanded of the students in class. They also seek reasonably common criteria for grading. Thus, a testing program would have instructors closely involved in testing

> for the specific skills needed to succeed in a given class as identified by analyzing lectures, instructional materials, and assignments. Such a process has not been popularly adopted. (p. 109)

Cohen also noted that as long as the faculty "guard the criteria on which they award student grades and allow those criteria to shift according to the abilities of the students in their classes, correlations between grades and any type of entrance test must remain low." (p. 109). Thus there will always be some degree of variation in instructional standards and grading policy. Short of comprehensive exit exams that mirror the skills assessed on placement tests, there will always be a relatively low point-to-point correspondence between final grade, retention, and placement test scores. The question then focuses on how much instructor grade point variation is tolerable? There are no references in the literature to give us the answer. Each college or department will have to debate the issue.

One guideline that emerges in the literature on inter-rater reliability is that the range of instructor GPA's should not exceed one-half a grade. Whatever range becomes the accepted norm, there will have to be a procedure to track the allocation of grades by instructor. It should be noted that an unusually high instructor GPA does not automatically indicate a problem with that instructor. Perhaps the teacher is exemplary and uses mastery teaching so that nearly all students finish the course with an A grade. High GPA variation within a department, particularly those that mandate minimum cutting scores for entry are only a possible signal to a problem.

Changing the method and approaches to placement testing in the community colleges to include multiple sources of dispositional data may require additional costs and resources. Placement test answer sheets may have to be modified to include questionnaire items relevant to the dispositional characteristics of the student. Colleges and testing bureaus may have to find ways to integrate dispositional data with test scores to yield a more complete profile of the matriculating student. Paper and pencil placement tests are

much in favor because of their relatively low cost, ease of administration, and the provision of a metric (test score) that ranks sorts and classifies students according to some inferred ability. Adding dispositional variables to the composite model for placement appears to provide substantial improvements to the course grade and retention prediction. Although a clear recommendation from this investigation about how to integrate dispositional data with test score information did not emerge it is strongly recommended that all institutions using a placement score incorporate multiple data points with respect to dispositional information.

There will be costs involved in an improved placement scheme that integrates more dispositional data about the student. However, the benefits may outweigh the costs. For example, a misplaced student results in opportunity costs for both the student and the institution. The opportunity cost for the student may be the time spent trying to master material that they were not ready for, forcing them to drop out or repeat the course. Conversely, an opportunity cost may come in the form of a student who was misclassified and prevented from enrolling in a course that they could have succeeded in. The student then has to extend the amount of time in training or education in courses that may not be necessary. Opportunity could be reduced if teaching methods were altered to better integrate the teaching of basic skills with academic or vocational skills (Sticht, et. al, 1987). The costs to the institution of including dispositional information may be substantial, particularly if institutions insist on using verifiable information such as data available form transcripts. However, this notion of cost and benefit separates facts from values. An espoused value of the community colleges and the matriculation program is to conduct assessment and placement in such a way that student success and access are made optimal. Making optimal and fair placement decisions for students ought to guide the policy and practice of selecting, classifying, and placing students in the open-access colleges.

Delimitations and Limitations

This investigation confined itself to first-time, freshman students matriculating at three large-sized community colleges located in urban southern California. Thus the findings produced by this investigation may not be generally applicable to other community colleges or other education and training institutions.

Students included in this study sat for an assessment test between May and September, 1994 and subsequently enrolled in a pre-collegiate or collegiate level English or mathematics course in the fall, 1994 term. Therefore this study did not include students not taking an assessment test, or who did not enroll in at least one of these two subject areas. Thus the findings from this study may not necessarily apply to students who apply to college but delay enrollment in college or in English and mathematics courses. Students who may have been discouraged by some aspect of the placement process and did not enroll in one of these two subject areas will not be included. Also, community college students often attend part-time, thus scheduling difficulties may preclude some from enrolling in a particular course, regardless of their eligibility for the course. However the study design did not include this group of students. This may have introduced some selection bias into the sample.

Only students without prior college experience were included. The use of first-time freshman students was intended to limit the effects of prior college experience on course performance and to focus more directly on the relation between assessment test scores, student characteristics, and course performance outcomes. Not all students in community colleges are first time students without prior college experience, thus these findings also may have introduced some selection bias into the sample and potentially limits the general application of these findings.

There are some limitations with respect to the interpretations and inferences produced from this investigation. First, this study is retrospective. It used testing, enrollment, and course outcome data from the fall, 1994 semester. It used data from an assessment and placement system that had been in use for several years. There was not the opportunity to randomly select students and assign them to treatments. Thus true experimental design with random assignment to groups was therefore not possible given the requirements of the study.

The majority of the demographic, situational, and dispositional data were gathered using a questionnaire administered to students. Therefore these data are self-reported and may over- or under-estimate the true characteristics of the student. None of the information reported by the student was verified, thus it is possible that the self reported data were not completely accurate with respect to depicting the biographical characteristics of the participants. However, the situational and dispositional data yielded generally statistically significant explanatory power in the course outcomes model. It has been noted in the literature that self-reported data are generally reliable and reflective of the true characteristics of the study participants (Miller, 1991).

Another limitation to the study is the possible truncation of range due to analysis of a pre-existing assessment and placement system. It has been noted by other investigators that low or moderate predictive validity coefficients do not necessarily mean that the measures used to predict performance are not related to the dependent variable of final grade (Anastasi, 1964; Cronbach, 1971).The colleges included in this investigation have used placement tests since 1986. For the first six years of the placement program in these three colleges, students were prevented from directly enrolling in a particular course unless they achieved the minimum cutting score on placement tests. However, starting in 1992, these mandatory placement rules were temporarily suspended when the colleges

implemented a new electronic registration system. Students were then able to ignore placement advice and enroll in any level of English or mathematics course. Relaxation of placement rules may reduce a major confound to predictive validity studies. For college in which a placement program has been implemented, it is likely that many students followed the course recommendation given by the college. However, the ability of students to ignore placement advice and attempt a course of their choosing increases the range student within a course. This temporary suspension in mandatory placement rules reduces, but does not eliminate, the possible truncation of score ranges within each level of the English or mathematics curriculum. The application of a correction for a restriction of range was used to analyze the reliability of the predictive validity coefficients between placement tests and final grades. This was an attempt to mitigate this threat to the validity of the findings.

Technical constraints in the measures of final grade, retention, and placement test scores may have artificially weakened or lowered the validity coefficients. These constraints include restriction of range in the criterion variable of final grade, restriction of range on the predictor as described in the preceding paragraph, and instability in the criterion.

Directions for Future Research

This investigation provides guidance to other researchers interested in improving or better understanding predictive validity. Predictive validity is essentially an attempt to approximate the future in the present. Institutions seeking to identify individuals with the best aptitude for success in a course are in essence attempting to predict how a person will perform at some point in the future under different conditions. Thus one recommendation might be for colleges to use the content and conditions of the course to analyze how well a student is likely to do or be retained in this setting. Although it may be difficult to

reproduce classroom conditions, students may be given examples of the type of literacy skills demanded in the course. At the point of assessment, registration, or counseling, samples of writing or mathematical problems might be available to show prospective students. These work samples might be constructed to reflect the types of English or mathematics skills used in various courses. Students might be assessed using a pre-test developed by the instructional faculty that better reflect local conditions and course objectives than a nationally standardized test. Reproducing for the prospective student the actual demands and requirements of the class would serve to better inform the student of the classroom demands, and provide the faculty with students who possess the desired skills and traits for success. For future research, a group of students might be referred using a standardized assessment instrument, while another group is referred using the locally developed assessment and informed decision process. The course grade and retention outcomes of the students might be compared to determine if more accurate and equitable placement resulted.

Another possible avenue of study might be to administer a teaching style inventory to a group of instructors and a learning style inventory to a group of students. Students could then be matched on the basis of the degree of fit between their particular learning style, and the instructional style of the faculty member. Outcomes from this experiment could be compared with students assigned to courses under the existing system.

This study did not focus on student sub-groupings. With the increasing ethnic and linguistic diversity of students coming to the community colleges, a study examining the differential placement and success rates of different groupings students could be conducted. A future study might analyze these same data for various student sub-groups by race, age, or sex to determine if there are interaction effects of dispositional, situational, and test score performance with instructor practices.

The assumption that the variance in standardized placement test scores is reflective of the varying abilities of students is only a very small part of the placement puzzle. This study suggests that test scores alone do not reduce variance in student background characteristics. Variance is also found in the various courses with respect to grading standards and retention practices. There are many pieces to the puzzle available to college administrators, faculty, and staff. These pieces, properly arranged may result in a better fit between the student's dispositional, situational characteristics, and instructional needs, and the instructor's approach to delivering instruction. Arranging a more conducive fit between student needs and instructor characteristics may be one of the more difficult jobs facing selection and education personnel. The reward of this effort may be improved student course performance and retention.

The power of dispositional variables in explaining student course outcomes was demonstrated by this study. Dispositional variables have shown over the years to be often stronger predictors than test scores or other standardized measures of academic achievement. For example Cohen and Brawer (1987) found that student self assessment of ability in comparison with other students at their college was highly related to the content knowledge measured on the General Academic Assessment (GAA). Students were asked to compare themselves with other students at their college in their ability to "edit written material," or "use algebra to solve problems." The correlations between the student self assessments and the scales obtained from the content areas of the GAA were extremely high. This finding again suggests the power of dispositional data for inferring student ability. As noted by Cohen and Brawer, "Students know what they know" (p. 123).

The value of dispositional data in student placement might be demonstrated in a research study. For example, one avenue of research might use the predictive power of dispositional and situational variables to place students in lieu of standardized placement

test scores. A research study could place students using only self-reported dispositional data such as high school grade point average or the importance of college to the student. Another group of students could be placed using the cutting scores on a placement test. The two groups of students could then be compared on the basis of course performance and retention. If it can be shown that self-reported dispositional and situational data explain as much or greater amounts of variance in final grade or retention, then colleges may want to consider using a questionnaire to place students. In this way the cost, time, and scheduling of testing sessions and the attendant problems of test security, scanning, and compliance with legal statutes might be avoided. This may result in considerable cost savings to the institution, and may improve the accuracy and fairness of the inferences made about student aptitude. This would be particularly useful if, as this investigation suggests, that student biographical characteristics can place individuals with improved accuracy, even with the confound of substantial variation in grading practices and retention.

In summary, the present investigation revealed that of all the variables used to explain variance in student course performance, dispositional variables tended to account for greater variance in course outcomes of retention and final grade than did other biographical data, including placement test scores. Instructor grading variation was also found to significantly reduce error in the model indicating that grading inconsistency or differences in retention practices has a strong effect on student course performance. With respect to the employment status of the instructor, part-time faculty demonstrated greater variation in grading practices and retention outcomes than did full-time faculty.

Colleges may also want to pay careful attention to the effects of denying access to courses based on an inferred ability level or prerequisite. The repeated finding that ineligible students were generally 70%-80% as successful as eligible students suggests strongly that student access may be unfairly denied and that many students capable of

success in the course are not given a chance to try. Inaccurate referral to lower level courses increases the opportunity costs for the students, and misallocates the educational resources of the college. Many community college students must attend part-time and have family responsibilities. Extending their time in training or education unnecessarily may add greatly to the costs of attending school, and delay student transfer to the university or entry into and promotion in the job market. Colleges implementing mandatory as opposed to advisory prerequisites might consider the equity and opportunity costs of this policy, particularly for students incorrectly classified as probable failures.

The importance of the study lies in the finding that student placement can be enhanced by using self-reported dispositional data rather than strict reliance on cutting scores on a standardized test for course eligibility. The present investigation also highlighted the importance of educational standards in the community college. Standards must be viewed as imposed on the student by the college, and not what the students bring to the college. Higher selectivity does not necessarily improve or make academic standards consistent.

APPENDIX 1

Placement Test Scores by Demographic Characteristics of Student Sample

Significant Variables for English and

Mathematics Course Grade and Retention Models

Table A-1: Placement Test Scores by Demographic Characteristics of Participant Sample

Group	Reading Placement Test		Writing Placement Test		Mathematics Placement Test	
	Mean	Std. Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Sex						
Male	18.5	7.3	21.5	7.2	26.2	10.1
Female	17.2	7.2	21.3	7.1	23.9	9.2
Significance*	***		NS		***	
Race						
American Indian	20.7	6.9	23.4	6.7	26.1	9.5
Asian	13.9	6.3	18.1	6.8	22.4	9.7
African American	16.4	6.4	19.0	6.0	21.1	8.8
White	21.8	6.6	25.2	6.6	28.6	9.4
Latino-Hispanic	15.8	6.5	19.8	6.2	23.0	9.2
Filipino	18.4	6.5	22.4	5.9	26.2	8.5
Other (non-white)	14.9	7.4	19.0	7.9	20.1	9.1
Significance*	***		***		***	
English Primary Language						
Yes	20.2	6.7	23.5	6.6	26.0	9.6
No	12.8	5.6	17.3	6.2	20.1	8.3
Significance*	***		***		***	
Verified Learning Disability						
Yes	16.0	7.3	17.9	6.5	19.9	9.9
No	18.0	7.2	21.9	6.9	25.0	9.5

* Significance based on chi-square statistic ***p<.001

Table A-2: Significant Variables as Measured by Standardized Beta Weights for English Course Grade Prediction Model

Variables		Beta	F	Probability
Placement Test				
	Writing Placement Test	.22	32.5	p<.05
Demographic	Race: African American	-.09	9.3	p<.05
	Sex: Female	.13	20.1	p<.05
Dispositional	High School GPA	.16	14.0	p<.05
	Grade in Last English Course	.14	6.0	p<.05
	Importance of College to You	.07	3.72	p<.05
	Non-Veteran	-.11	7.1	p<.05
Situational				
	Years Out of School	.14	5.3	p<.05
Instructor		***	3.8	p<.05

Table A-3: Significant Variables as Measured by Standardized Beta Weights for Mathematics Course Grade Prediction Model

Variables		Beta	F	Probability
Demographic	Race: African American	-.16	15.1	p<.05
	Race: Other	-.14	13.6	p<.05
	Sex: Female	.13	20.1	p<.05
	Age	.11	3.9	p<.05
Dispositional	High School GPA	.09	4.9	p<.05
	Highest Level Math Class	.19	19.1	p<.05
Situational	How Long Ago Last Math Class	.15	6.2	p<.05
Instructor		***	4.9	p<.05

Table A-4: Significant Variables for English Course Retention Model

Variables		Wald Statistic	Probability
Placement Test	Writing Placement Test	5.04	p<.05
Demographic	Race: Asian	4.33	p<.05
	Race: Latino	9.10	p<.05
	Race: White	5.41	p<.05
	Race: Filipino	13.37	p<.05
Dispositional	High School GPA	7.11	p<.05
Instructor		***	p<.05

Table A-5: Significant Variables for Mathematics Course Retention Model

Variables		Wald Statistic	Probability
Demographic	English Not Primary Language	8.51	p<.05
Dispositional	High School GPA	8.91	p<.05
	Years of English Completed	3.99	p<.05
	Attended Above Standard High School	13.04	p<.05
	Non-Veteran	4.90	p<.05
Situational	Evening Student	4.22	p<.05
Instructor Block		***	p<.05

APPENDIX 2

Copy of Questionnaire

TODAY'S DATE		
MONTH	DAY	YEAR

PLEASE PRINT TODAY'S DATE

DIRECTIONS

- USE NO. 2 PENCIL ONLY
- Fill in bubble completely Ⓐ Ⓑ ●
- Erase completely to change

LAST NAME | **FIRST NAME** | **M.I.**

SOCIAL SECURITY NUMBER

DATE OF BIRTH

MONTH	DAY	YEAR

IDENTIFICATION NUMBER (If required)

1. ETHNIC BACKGROUND
(Choose only one)

- ○ American Indian/Alaskan Native
- ○ Asian
- ○ Pacific Islander
- ○ Black/not of Hispanic Origin
- ○ White/not of Hispanic Origin
- ○ Hispanic
- ○ Filipino
- ○ Other

2. SEX

○ Male ○ Female

3. IS ENGLISH YOUR FIRST (PRIMARY) LANGUAGE?

○ Yes ○ No

4. DO YOU HAVE A VERIFIED LEARNING DISABILITY (SUCH AS DYSLEXIA)?

○ Yes ○ No

5. WHAT IS YOUR ADMISSION STATUS AT THIS COLLEGE?

- ○ New, first time in any college.
- ○ New to this college but have attended other college
- ○ Returning to this college after absence from this college
- ○ Continuing at this college

6. ARE YOU A VETERAN?

○ Yes ○ No ○ Active Duty

FOR OFFICE USE ONLY

SCORE 1	SCORE 2	SCORE 3	SCORE 4	SCORE 5	READING RATE (WPM)

CAPP

Computerized Assessment and Placement Programs

Form No. F-574-CAPP

Mark Reflex® by NCS MM206917-1 321 EO11 Printed in U.S.A.

041177 MAKE NO MARKS IN THIS AREA

EDUCATIONAL BACKGROUND

7. HIGH SCHOOL EDUCATION: *(Choose only one)*

- ○ Still in H.S.
- ○ Not a H.S. graduate
- ○ High School diploma
- ○ Foreign secondary diploma
- ○ G.E.D.
- ○ H.S. proficiency
- ○ Certificate of completion

8. HIGHEST COLLEGE DEGREE OR CERTIFICATE EARNED:

- ○ No degree at this time
- ○ Certificate
- ○ Associate Degree
- ○ Bachelor's Degree
- ○ Master's Degree or beyond
- ○ Other

9. HOW LONG HAVE YOU BEEN OUT OF SCHOOL? *(Don't count or include summer)*

- ○ Still in school
- ○ Less than 1 year
- ○ 1-2 years
- ○ 3-4 years
- ○ 5-10 years
- ○ More than 10 years

10. HOW MANY YEARS OF ENGLISH HAVE YOU COMPLETED IN HIGH SCHOOL? *(Do not include ESL)*

- ○ Less than 1 year in high school
- ○ 1 year in high school
- ○ 2 years in high school
- ○ 3 years in high school
- ○ 4 years in high school

11. WHAT GRADE DID YOU RECEIVE IN THE LAST ENGLISH CLASS YOU COMPLETED?

Ⓐ Ⓑ Ⓒ Ⓓ Ⓕ

12. WHAT IS YOUR HIGH SCHOOL GRADE POINT AVERAGE *(G.P.A.)?*

- ○ A- to A 3.5–4.0
- ○ B to A- 3.0–3.4
- ○ B- to B 2.5–2.9
- ○ C to B- 2.0–2.4
- ○ C- to C 1.5–1.9
- ○ D to C- 1.0–1.4
- ○ Below D 0.–0.9

13. HIGH SCHOOL LAST ATTENDED: *(Use code sheet)*

⓪	⓪	⓪	⓪
①	①	①	①
②	②	②	②
③	③	③	③
④	④	④	④
⑤	⑤	⑤	⑤
⑥	⑥	⑥	⑥
⑦	⑦	⑦	⑦
⑧	⑧	⑧	⑧
⑨	⑨	⑨	⑨

14. WHAT IS THE HIGHEST LEVEL MATH CLASS YOU HAVE COMPLETED?

- ○ None
- ○ Basic math (arithmetic)
- ○ Algebra I (beginning)
- ○ Geometry
- ○ Algebra II (intermediate)
- ○ Trigonometry
- ○ College Algebra/Pre-calculus
- ○ Calculus

15. WHAT GRADE DID YOU RECEIVE IN THE LAST MATH CLASS YOU COMPLETED?

Ⓐ Ⓑ Ⓒ Ⓓ Ⓕ

16. HOW LONG AGO DID YOU COMPLETE YOUR LAST MATH CLASS?

- ○ Currently enrolled in a math course
- ○ Less than 1 year
- ○ 1-2 years
- ○ 3-5 years
- ○ More than 5 years

COLLEGE PLANS

17. PLAN TO ATTEND:

○ Day ○ Evening ○ Day and evening

18. COLLEGE UNITS PLANNED FOR NEXT TERM:

- ○ Less than 6 units
- ○ 6-8 units
- ○ 9-11 units
- ○ 12 units or more

19. EMPLOYMENT HOURS PLANNED WHILE ENROLLED:

- ○ None
- ○ 1-10 hours/week
- ○ 11-20 hours/week
- ○ 21-30 hours/week
- ○ 31-40 hours/week
- ○ More than 40 hours/week

20. MOST IMPORTANT EDUCATIONAL GOAL:

- ○ Personal enrichment only
- ○ H.S. diploma
- ○ Vocational training/ certificate
- ○ Two year degree
- ○ Transfer to a 4 year college
- ○ Other

21. I WOULD LIKE INFORMATION ABOUT: *(Mark all that apply)*

- ○ Financial aid
- ○ Job placement
- ○ Child care
- ○ Disabled student services
- ○ Health services
- ○ Learning disabilities
- ○ Veterans' program
- ○ E.O.P.S.
- ○ Re-entry program
- ○ Work experience
- ○ Orientation to college
- ○ Choosing a major
- ○ Career planning
- ○ Counseling
- ○ Transferring
- ○ Tutoring
- ○ Study skills
- ○ Math skills
- ○ Reading skills
- ○ Writing skills
- ○ E.S.L.

22. MAJOR: Use code sheet

⓪	⓪	⓪	⓪
①	①	①	①
②	②	②	②
③	③	③	③
④	④	④	④
⑤	⑤	⑤	⑤
⑥	⑥	⑥	⑥
⑦	⑦	⑦	⑦
⑧	⑧	⑧	⑧
⑨	⑨	⑨	⑨

23. HOW DEFINITE IS YOUR CHOICE OF MAJOR?

- ○ Very
- ○ Fairly
- ○ Unsure

24. TRANSFER COLLEGE:

⓪	⓪	⓪	⓪
①	①	①	①
②	②	②	②
③	③	③	③
④	④	④	④
⑤	⑤	⑤	⑤
⑥	⑥	⑥	⑥
⑦	⑦	⑦	⑦
⑧	⑧	⑧	⑧
⑨	⑨	⑨	⑨

25. HOW IMPORTANT IS IT TO THE PEOPLE CLOSEST TO YOU THAT YOU GO TO COLLEGE?

- ○ Not very important
- ○ Somewhat important
- ○ Very important

26. HOW IMPORTANT IS COLLEGE TO YOU PERSONALLY?

- ○ Not very important
- ○ Somewhat important
- ○ Very important

27. ADDITIONAL QUESTIONS: *Optional*

1.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
2.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
3.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
4.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
5.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
6.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
7.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
8.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
9.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
10.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
11.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
12.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
13.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
14.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
15.	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ

REFERENCES

Alkin, M.C. and Freeman, M. (1991). Matriculation evaluation: Summary report 1990-1991. Northridge, CA: Educational Evaluation Associates.

Alkin, M.C. and Freeman, M. (1992). Matriculation evaluation: Summary report 1991-1992. Northridge, CA: Educational Evaluation Associates.

American Psychological Association. (1985). Standards for educational and psychological testing. Washington, DC: Author.

Anastasi, A. (1968) Psychological testing (3rd ed.) New York: Macmillan

Armstrong, W.B. (1989). Honors program evaluation report. San Diego, CA: San Diego Community College District, Institutional Research Office. (ERIC Document Reproduction Service Number No. ED 313 066)

Armstrong, W.B. (1995). English placement testing, multiple measures and disproportionate impact: An analysis of the criterion- and content-related validity evidence. San Diego, CA: Office of Institutional Research, San Diego Community College District. (ERIC Document Reproduction Service Number No. ED 398 965)

Armstrong, W.B. (1997). [Survey of California community colleges testing instruments and cutting scores]. Unpublished raw data. San Diego, CA: Office of Institutional Research, San Diego Community College District.

Assembly Bill 1725 (1986). Sacramento, CA: Legislative Analyst Office.

Asher J. J. and Sciarrino, J.A. (1974). Realistic work sample tests: A review. Personnel Psychology, 27, 519-533.

Asher, J.J. (1972). The biographical item: Can it be improved? Personnel Psychology, 25, 251-269.

Astin, A. W. (1991). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education. New York, Macmillan

Astin, A. W. (1985). Achieving educational excellence: A critical assessment of priorities and practices in higher education. San Francisco, CA: Jossey-Bass.

Astin, A.W. (1975). Preventing students from dropping out. San Francisco: Jossey-Bass

Bean, J.P. (1982). Conceptual models of student attrition: How theory can help the institutional researcher. In E. T. Pascarella (Ed.). New Directions for Institutional Research, 36, 17-33; San Francisco: Jossey-Bass.

Bean, J.P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. Research in Higher Education, 12, 155-187.

Beatty-Guenter (1992). Sorting, supporting, connecting, and transforming: Student retention strategies at community colleges. (ERIC Document Reproduction Service No. ED 342 425)

Berger, D.M. (1997). Mandatory assessment and placement: The view from an English department." Appearing in J.M. Ignash (Ed.). New Directions for Community Colleges, 100, 33-42. San Francisco, CA: Jossey Bass.

Bers, T. (1989). Assessing assessment programs: The theory and practice of examining reliability and validity of a writing placement test. Community College Review, 18 17-27.

Bers, T. (1997). Self assessments, academic skills, and student achievement. Journal of Applied Research in the Community College. 4 101-118.

Beusse, W. (1974). The impact of military service on low-aptitude men. (AFHRL-TR-74-74.) Brooks AFB, TX: Air Force Human Resources Laboratory.

Birenbaum, W. M. (1986). From mass to class in higher education. Appearing in L.S. Zwerling (Ed.). New Directions for Community Colleges, 54, 3-12 San Francisco: Jossey-Bass.

Boggs, G. (1984). The effect of basic skills assessment on student achievement and persistence at Butte College: A research report. (ERIC Document Reproduction Service No. ED 244 686)

Bogue, Jesse P. (1950). The community college. (1st ed.). New York: McGraw-Hill.

Boughan, Karl (1995). Developmental placement and academic progress: Tracking "At-Risk" students in the 1990 entering cohort. (Enrollment Analysis EA 95-2) Largo, MD: Prince George's Community College, Office of Institutional Research and Analysis. (ERIC Document Reproduction Service No. ED 377 908)

Bray, Dorothy (1987). Assessment and placement of developmental and high-risk students. In K. M. Ahrendt (Ed.): Teaching the developmental education student. New Directions for Community Colleges, 57 33-48. San Francisco, CA: Jossey-Bass.

Breneman, D. W., and Nelson, S. C (1981). Financing community colleges: An economic perspective. Washington, D.C.: Brookings Institution.

Brint, S. and Karabel. J. (1989). The diverted dream: Community colleges and the promise of educational opportunity in America, 1900-1985. New York: Oxford University Press.

Brownstein, R. B. (1997, August 11). Call for academic standards could face test from civil rights law. The Los Angeles Times, p. A5.

Cage, M. C. (1991, June 12). California community colleges system agrees to change role of testing. The Chronicle of Higher Education, p. A21.

California Code of Regulations (1990). Title 5, Statutes of 1990. Sacramento, CA: California State Printing Office.

California Community Colleges (1992). Standards, policies, and procedures for the evaluation of assessment instruments used in the California community colleges. Sacramento, CA: Matriculation Assessment Work Group, California Community Colleges.

Chabot College (1996). Prerequisite validation model. Author.

Clark, B.R. (1962). Educating the expert society. San Francisco: Chandler.

Cohen, A.M. and Associates (1970). Objectives for college courses. Beverly Hills, CA: Glencoe Press.

Cohen, A. M. (1975). College responses to community demands. San Francisco: Jossey-Bass.

Cohen, A.M. (1980). Dateline '79 revisited. In G. Vaughan, (Ed.), Questioning the community college role. New Directions for Community Colleges, 32 San Francisco: Jossey-Bass.

Cohen, A. M. (1985). Helping ensure the right to succeed: An ERIC review. Community College Review, 3 4-9.

Cohen, A.M. & F. Brawer (1987). The collegiate function of community colleges. San Francisco, CA: Jossey-Bass.

Cohen, A.M. & F. Brawer (1996). The American community college. (3rd ed.). San Francisco: Jossey-Bass.

College of the Canyons (1994). Predictive validity of the APS writing and reading tests and validating placement rules for the APS writing test. Santa Clarita, CA: College of the Canyons, Office of Institutional Development. (ERIC Document Reproduction Service No. ED 376 915).

Collier, M. (1986). A specific investigation of relative performance of examination markers. Assessment and Evaluation in Higher Education; 11 130-137. Ann Arbor, MI: University Microfilms International.

Collins, R. (1979). The credential society: An historical sociology of education and stratification. New York: Academic Press

Colvin, R. L. (1997, October 1). Should UC do away with the SAT? The Los Angeles Times, p B2.

Cronbach, L.J. (1990). Essentials of psychological testing (5th ed.). New York: Harper Collins Publishers.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (Ed.). Educational measurement, (2nd edition) Washington, DC: American Council on Education.

Cronbach, L.J. (1970). Essentials of psychological testing, (3rd ed.). New York: Harper & Row Publishers.

Cronbach, L.J., and G.Gleser (1965). Psychological tests and personnel decisions. (2nd ed.) Urbana, IL: University of Illinois Press.

Cronbach, L.J. and R.E. Snow (1969). Final report: Individual differences in learning ability as a function of instructional variables. Palo Alto, CA: School of Education, Stanford University.

Cross, K. P. (1985) Determining missions and priorities for the fifth generation. In Deegan, Tillery, & Associates (Ed.). Renewing the American community college. (pp. 34-50) San Francisco, CA: Jossey-Bass.

Cross, K.P. (1986, Spring). Excerpt from a speech delivered at San Diego City College, San Diego, CA: San Diego Community College District.

Cross, K.P. (1981). Adults as learners: Increasing participation and facilitating learning. San Francisco, CA: Jossey-Bass.

Deegan,W.L. and Tillery D. (1985). Renewing the American community college: Priorities and strategies for effective leadership. San Francisco: Jossey-Bass.

DeYoung, A.J. (1989). Economics and American education: A historical and critical overview of the impact of economic theories on schooling in the United States. New York: Longman.

Dorn, S. (1997). The political legacy of school accountability systems. The Education Policy Analysis Archives, 6, No 1.

Druckman, D. and Bjork, R.A. (Ed.) (1994). Learning, remembering, believing: Enhancing human performance.. Washington, D.C.: Commission on Behavioral and Social Sciences and Education, National Research Council: National Academy of Sciences, National Academy Press.

Dubin, R. and Taveggia, T.C. (1968). The teaching-learning paradox: A comparative analysis of college teaching methods. Eugene, OR: Center for the Advanced Study of Educational Administration, University of Oregon.

Education Commission of the States (1980). ECS law and education center footnotes No. 2. Denver, CO: Author.

Elmers, M.T. and G.R. Pike (1997). Minority and nonminority adjustment to college: Differences or similarities?. Research in Higher Education, 38 77-97.

Eitleberg, M., Laurence, J.H., & Brown, D. C. (1992). Becoming brass: Issues in the testing recruiting, and selection of American military officers. In B. R. Gifford (Ed.), Test policy in defense: Lessons from the military for education, training, and employment. (pp. 1-78). Boston, MA: Kluwer Academic Publishers.

Eitleberg, M., J. Laurence, B. Waters, and L. Perelman (1984). Screening for service: Aptitude and education criteria for military entry. Alexandria, VA: Human Resources Research Organization.

Fields, C (1988, June 1). "California lawsuit challenges 2-year colleges' practice that restricts low test scorers to remedial courses." The Chronicle of Higher Education, p. A1

Fonte, R. (1997). Structured versus laissez-faire open access: Implementation of a proactive strategy. New Directions for Community Colleges, 100, 43-52. San Francisco, CA: Jossey Bass.

Freeman, M.E. & Alkin, M.C. (1996). Matriculation evaluation: Technical assistance visits summary report 1995-1996. Sacramento, CA: State Chancellor's Office of the California Community Colleges.

Freeman, Richard B. (1976). The over educated American. New York: Academic Press.

Gabe, L.C. (1989). Relating college-level course performance to ASSET placement scores. (Institutional Research Report Abstract RR89-22). Fort Lauderdale, FL: Broward Community College. ERIC Document Reproduction Service No. 309 823).

Gifford B. R. (ed.) (1989). Test policy and test performance: Education, language, and culture. National Commission on Testing and Public Policy. Boston, MA: Kluwer Academic Publishers.

Gillespie, M. (1993). Placement testing in community colleges: A response to Hughes and Nelson. Community College Review, 20, (4), 59-69.

Glazer, N. (1970). Are academic standards obsolete? Change, 2 (6), 38-44.

Goodlad, J (1969). cited in G. Heathers Grouping In R.L. Ebel (Ed.). Encyclopedia of Educational Research, 4th ed. London: The Macmillan Co.

Hargis, C. H. (1990). Grades and grading practices: Obstacles to improving education and to helping at-risk students. Springfield, IL: Charles C. Thomas Publishing

Hendrick, L. (1980). Community education: Retrench or redesign? Community College Review, 7, (4), 51-57

Henriksen, J.A. (1995). The influence of race and ethnicity on access to postsecondary education and the college experience. An ERIC Digest. (ERIC Document Reproduction Service No. ED 386 242).

Herrnstein, R. J. and C. Murray (1994). The bell curve: Intelligence and class structure in American life. New York: The Free Press.

Hewitt, B. & R. Jacobs (1978). Student perceptions of grading practices in different major fields Journal of Educational Measurement, 15 213-218.

Holland, J. L. & Astin, A. W. (1962). The prediction of the academic, artistic, scientific, and social achievement of undergraduates of superior scholastic aptitude. Journal of Educational Psychology, 53, 132-143.

Hosmer, D.W. and S. Lemeshow (1989). Applied logistic regression. New York: John Wiley and Sons.

Hoyt, D.P. (1965). The relationship between college grades and adult achievement-a review of the literature. (Research Report Number 7). Iowa City, IA: American College Testing Program (ERIC Document Reproduction Service No. ED 023 343).

Hudson, J. Blaine, McPhee, S.A., and Petrosko, J. (1993). The relationship between tests, course placement, and the academic performance of college freshmen. NACADA Journal: 13 (2), 5-14.

Hughes, R.E. and C.H. Nelson (1991). Placement scores and placement practices: An empirical analysis. Community College Review, 19 (1), 42-48.

Ignash, J.M. (1997). Who should provide postsecondary remedial developmental education? Appearing in J.M., Ignash, (Ed.), New Directions for Community Colleges. 100, pp. 5-19. San Francisco, CA: Jossey-Bass.

Isonio, S. (1995). Prerequisite validation standards at Golden West College. Paper presented at the meeting of the Research and Planning Group for the California Community Colleges. Pasadena, CA: Pasadena City College.

Isonio, S. (1993). Implementation and initial validation of the Combined English Language Skills Assessment (CELSA) at Golden West College. Huntington Beach, CA: Golden West College (ERIC Document Reproduction Service No. ED 353 023).

Isonio, S. (1992). Implementation and initial validation of the MDTP tests at Golden West College. Huntington Beach, CA: Golden West College. (ERIC Document Reproduction Service No. ED 345 782).

Isonio, S. (1991). Implementation and initial validation of the APS English-writing test at Golden West College: Evidence for predictive validity. . Huntington Beach, CA: Golden West College. (ERIC Document Reproduction Service No. ED 345 781).

Jackson, G. (1990) appearing in T.G. Sticht (Ed.), Testing and assessment in adult basic education and English as a second language programs. Washington DC: US Department of Education, Division of Adult Education and Literacy.

Jansson, B.S. (1988). The reluctant welfare state: A history of American social welfare policies. Belmont, CA: Wadsworth Publishing.

Jencks, C. (1989). If not tests, then what?: Conference remarks. In B. R. Gifford (Ed.), Test Policy and Test Performance: Education, Language, and Culture. (pp. 115-122). Boston, MA: Kluwer Academic Publishers.

Jenkins, B. G. (1991). Meeting the diverse needs of two-year college students through appropriate course placement. Paper presented at the Fall Conference of the North Texas Community/Junior College Consortium. Dallas, TX.

Jensen, A.R. (1993). Psychometric g and achievement. In B. R. Gifford (Ed.), Policy Perspectives on Educational Testing. (pp. 117-227). Boston, MA: Kluwer Academic Publishers.

Jones, S. (1997 September 27). Ranking of city's schools rankles many. The San Diego Union, p. B1.

Kaplan, R.M. and Sacuzzo, D.P. (1982). Psychological testing: Principles, applications, and issues. 3rd edition. Pacific Grove, CA: Brooks/Cole Publishing Company.

Karabel, J. (1986). "Community colleges and social stratification in the 1980's." In L. S. Zwerling (Ed.), The community college and its critics. New Directions for Community Colleges, 54. San Francisco: Jossey-Bass.

Keller, D., Crouse, J., and Trusheim, D. (1994). The effects of college grade adjustments on the predictive validity and utility of SAT scores. Research in Higher Education, 35 2 195-208

Kingan, M.E. and R.L. Alfred (1993). Entry assessment in community colleges: Tracking or facilitating? Community College Review, 21 3 3-16.

Kirst, M. W. (1997, Winter). Exam confusion: Admissions and placement tests lack standardization. Crosstalk. San Jose, CA: The California Higher Education Policy Center.

Knoell, D. (1980). Planning in community colleges. In P. Jedamus, M.W. Peterson, and Associates (Ed.), Improving academic management: A handbook of planning and institutional research. San Francisco: Jossey-Bass.

Lee, V.E and Frank, K. (1989). Student characteristics which facilitate transfer from 2-year to 4-year colleges. Ann Arbor, MI: University of Michigan. (ERIC Document Reproduction Service No. ED 315 124)

Lemann, N. (1995, August). The structure of success in America. The Atlantic Monthly, 276, 2, pp. 41-60

Lemann, N. (1995, September). The great sorting. The Atlantic Monthly, 267 3 pp. 84-100.

Lenning, O.T. (1982). Variable selection and measurement concerns. In E. Pascarella (Ed.), New Directions for Institutional Research: Studying Student Attrition. No. 36. pp. 35-54. San Francisco, CA: Jossey-Bass.

Lewis, W.A., G. Dexter, and W.C. Smith (1978). Grading practices and test validation: A proposed new approach. Journal of Educational Measurement, 15 219-227

Linn, R.L. (Ed.). (1989). Educational measurement. New York: American Council on Education Macmillan.

Linn, R.L. (1972). Some implications of the Griggs decision for test makers and users. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 073 128)

Manning, R. (1991). Suffolk Community College E.C.R.P. -- Eastern Campus Retention Program. Riverhead, NY: Suffolk County Community College, Eastern Campus. (ERIC Document Reproduction Service No. ED 346 915)

Mathematics Testing Diagnostic Project (1991). User manual. San Diego, CA: University of California, San Diego.

McCabe, R.H. (1989). Floridians at risk of losing higher education opportunity. Miami, FL: Miami-Dade Community College. (ERIC Document Reproduction Service No. 313 087)

Messick, S. (1989). appearing in R. Linn, (Ed.) Educational Measurement. New York: American Council on Education /Macmillan (1989).

Miller, D.C. (1991). Handbook of research design and social measurement. 5th ed. Newbury Park, CA: Sage Publications.

Miner, J.B. (1974). Psychological testing and fair employment practices: A testing program that does not discriminate. Personnel Psychology, 27 49-62

Miyazaki, Ichisada (1981). China's examination hell: the civil service examinations of Imperial China. Translated by Conrad Schirokauer. New Haven: Yale University Press.

Montmarquette, C. and Mahseredjian, S. (1989). Could teacher grading practices account for unexplained variation in school achievements? Economics of Education Review: 8 (4), 335-342. Elmsford, New York: Pergamon Press, Inc.

Morante, E.A. (1989). Selecting tests and placing students. Journal of Developmental Education, 13 2 2-6

Mount San Antonio College (1989). Course retention analysis focus study. Walnut, CA: Mount San Antonio College. (ERIC Document Reproduction Service No. ED 313 064)

National Center for Education Statistics (1996). Digest of education statistics: 1996. Washington DC: Office of Educational Research and Improvement, US Department of Education.

National Commission on Excellence in Education (1983). A nation at risk: The imperative for educational reform. A report to the nation and the secretary of education. Washington, D.C.: U.S. Government Printing Office.

National Commission on Testing and Public Policy (1990. From gatekeeper to gateway: Transforming testing in America. Chestnut Hill, MA: Boston College.

Norusis, M.J. (1994). SPSS advanced statistics 6.1. Chicago, IL: SPSS, Inc.

Oakes, Jeannie (1985). Keeping track: How schools structure inequality. New Haven: Yale University Press.

O' Connor, M.C. (1989). Aspects of differential performance by minorities on standardized tests: Linguistic and sociocultural factors. Appearing in (B. Gifford (Ed.)Test Policy and Test performance: Educational, Language, and Culture Boston, MA: Kluwer Academic Publishing. pp. 129-181

Orfield, G. & Paul, F. (1992). State higher education systems and college completion: Final report to the Ford Foundation. (ERIC Document Reproduction Service No. ED 354 4041)

Owen, D. (1985). None of the above: Behind the myth of scholastic aptitude. Boston, MA: Houghton-Mifflin

Pascarella, E.T. (Ed.) (1982). Studying student attrition. New Directions for Institutional Research, no. 36. San Francisco: Jossey-Bass.

Pascarella, E.T. & Terenzini, P.T. (1991). How college affects students: findings and insights from twenty years of research. San Francisco: Jossey-Bass.

Pedhazur, E.J. (1982). Multiple regression in behavioral research: Explanation and prediction. Fort Worth, TX: Holt, Rinehart and Winston.

Petrillo, J.E. (1971). Are your employment practices legal?. Junior College Research Review; (6) 3. Los Angeles, CA: ERIC Clearinghouse for Junior Colleges. ERIC Document Reproduction Service No. ED 054 768.

Pike, G. R. (1995). The relationship between self-reports of college experiences and achievement test scores. Research in Higher Education, 34 23-39.

Popham, W. J. (1990). Modern educational measurement: A practitioner's perspective 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Prus J. and R. Johnson (1994). A critical review of student assessment options. New Directions for Community Colleges, 88 69-84

Rasor, R.A. and J. Barr (1993). Refinement in assessment validation: Technicalities of dealing with low correlations and instructor grading variation. Sacramento, CA: American River College.

Reutter, E.E. (1975). Tests and employment discrimination. New York: Columbia University Institute of Administrative Research. ERIC Document Reproduction Service No. ED 107 733

Rogers, P. H. (1990). Student retention and attrition in college. In Robert M. Hashway. (Ed.) Handbook of developmental education (pp. 307-327). New York: Praeger.

Roueche, J.E. and S. D. Roueche (1993). Between a rock and a hard place. Washington, DC: American Association of Community Colleges.

Roueche, J.E., G.A. Baker, and S.D. Roueche (1984). College responses to low-achieving students: A national study. Orlando, FL: Harcourt Brace Johanovich Media Systems.

Roueche, J.E. and P.F. Archer (1979). Entry level assessment in college. Community College Review, 6, (4), pp. 15-27.

Roueche, J.E., Baer, G. A., and Brownell, R. L. (1971). Accountability and the community college: Directions for the 70's. Washington, DC: American Association of Junior Colleges.

Rucker, T. (1997, September 23). Texas changes to skills test cause confusion. Community College Times pp. 1, 7.

Schwarz, P.A. (1971). Prediction instruments for educational outcomes. In R.L.Thorndike, (Ed.). Educational measurement, 2nd edition. Washington, DC: American Council on Education.

Sheldon, M. S. (1970). Entrance and placement testing for the junior college. Junior College Research Review, Dec. 1970, 5, 1-4.

Shields, J. L., and F. C. Grafton (1983). A natural experiment: Analysis of an almost unselected army population. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Simon, J.L. and P. Burstein (1985). Basic research methods in social science. New York: Random House.

Sinaiko, W. (1988). Personal communication to Dr. Thomas G. Sticht.

Smittle, P. (1995). Academic performance predictors for community college student assessment. Community College Review 23 (2), pp. 37-46.

Spady, W. (1971). Dropouts from higher education: Toward an empirical model. Interchange 2 38-62.

State Chancellor's Office (1991). Assessment validation project: Local research options. Sacramento, CA: State Chancellor's Office of the California Community Colleges.

Sticht, T.G. (1992). Military testing and public policy: Selected studies of lower aptitude personnel. In B. R. Gifford (Ed.), Test policy in defense: Lessons from the military for education, training, and employment. (pp. 1-78). Boston, MA: Kluwer Academic Publishers.

Sticht, T. G. (ed.) (1975). Reading for working: A functional literacy anthology. Alexandria VA: Human Resources Research Organization.

Sticht, T. G., Armstrong, W.B., Hickey, D.T., and Caylor, J.S. (1986). Cast-off Youth: Policy and training methods from the military experience. Westport, CT: Praeger.

Sticht, T.G. (1990, January). Testing and assessment in adult basic education and English as a second language programs. Washington DC: US Department of Education , Division of Adult Education and Literacy.

The College Board (1984). Assessment and placement services for the community colleges: Using and interpreting scores. Princeton, NJ: College Entrance Examination Board.

The College Board (1989). Guide to the use of the descriptive tests of language skills. Princeton, NJ: College Entrance Examination Board.

Thompson, D.E. and Thompson, T.A. (1982). Court standards for job analysis in test validation. Personnel Psychology, 35, 865-874

Thorndike, E.L., and Woodworth R.S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. Psychological Review 8, 247-261.

Thorndike, R.L. (Ed.) (1971). Educational measurement, 2nd edition. Washington, DC: American Council on Education.

Thorndike, R.L. (1991). Measurement and evaluation in psychology and education. New York: Macmillam.

Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition. Chicago: University of Chicago Press.

Tinto, V. (1982). "Defining dropout: A matter of perspective." In E.T. Pascarella (Ed.), Studying Student Attrition. New Directions for Institutional Research, 9 4, pp. 3-16. San Francisco, CA: Jossey-Bass.

Tyler, Ralph W. & Wolf, Richard M. (1974). Critical issues in testing. Berkeley, CA: McCutchan Publishing.

U.S. Department of the Army (1965). Marginal man and military service. Washington, DC: U.S. Government Printing Office.

Wall, M. (1996). From theory to practice: Using retention research to guide assessment efforts at a community college. Mays Landing, NJ: Atlantic Community College. (ERIC Document Reproduction Service No. ED 397 929).

Warren, J. (1985). The changing characteristics of community college students. In W. Deegan and D. Tillery (Ed.), Renewing the American community college: Priorities and strategies for effective leadership. (pp. 53-79). San Francisco, CA: Jossey-Bass.

Wattenbarger, J. L. and McLeod, N. (1989). Placement in the mathematics curriculum: What are the keys? Community College Review, 16 (4), pp. 17-21.

Wenrich, J.W. (1995). Mission possible: Teaching the disenfranchised. Abstract of keynote address to the 1994 Leadership 2000 Conference. Mission Viejo, CA: League for Innovation in the Community College.

Western Association of Schools and Colleges (WASC) (1996). Handbook of accreditation and policy manual. 1996 edition. Santa Rosa, CA: Accrediting Commission for Community and Junior Colleges, Western Association of Schools and Colleges .

Willingham, W.W. (1974). College placement and exemption. New York: College Entrance Examination Board.

Yoakum, C.S. and Yerkes, R.M. (1920). Army mental tests. New York: H. Holt and Company.

Zwerling, L.S. (1986). Lifelong learning: A new form of tracking. In L.S. Zwerling (Ed.), The Community College and its Critics. New Directions for Community Colleges, 54. San Francisco, CA: Jossey-Bass.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Explaining Community College Outcomes by Analyzing Student Data and Instructor Effects

Author(s): William B. Armstrong

Corporate Source:

Publication Date: ~~Dec~~ Jan. 1999

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: William B Armstrong

Organization/Address:

Printed Name/Position/Title: William B Armstrong, Ph.D.

Telephone: 619-459-7128 FAX:

E-Mail Address: barmstro@sdccd.cc.ca.us Date: 3/24/99

(over)